A repeat protein links Rubisco to form the eukaryotic carbon concentrating organelle

Luke C. M. Mackinder[a], Moritz T. Meyer[b], Tabea Mettler-Altmann[c,1], Vivian Chen[a,d], Madeline C. Mitchell[b,2], Oliver Caspari[b], Elizabeth S. Freeman Rosenzweig[a,d], Leif Pallesen[a], Gregory Reeves[a,3], Alan Itakura[a,d], Robyn Roth[e], Frederik Sommer[c,4], Stefan Geimer[f], Timo Mühlhaus[c,4], Michael Schroda[c,4], Ursula Goodenough[e], Mark Stitt[c], Howard Griffiths[b], Martin C. Jonikas[a,5]

[a]Carnegie Institution for Science, Department of Plant Biology, Stanford, California 94305, USA. [b]Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK. [c]Max Planck Institute of Molecular Plant Physiology, 14476 Potsdam-Golm, Germany. [d]Department of Biology, Stanford University, Stanford, CA 94305, USA. [e]Department of Biology, Washington University, St. Louis, Missouri 63130, USA. [f]Cell Biology & Electron Microscopy, University of Bayreuth, 95440 Bayreuth, Germany.

Present addresses: [1]Cluster of Excellence in Plant Sciences and Institute of Plant Biochemistry, Heinrich-Heine-University, 40225 Düsseldorf, Germany; [2]CSIRO Agriculture, Canberra, ACT, Australia; [3]Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK; [4]Institute of Molecular Biotechnology and Systems Biology, Kaiserslautern University of Technology, 67653 Kaiserslautern, Germany.

[5]To whom correspondence should be addressed. Email: mjonikas@carnegiescience.edu. Telephone: +1 650 739 4216

Classification: Biological sciences: Cell Biology

Key Words: Carbon fixation, Rubisco, pyrenoid, *Chlamydomonas reinhardtii*, CO_2 concentrating mechanism

Biological carbon fixation is a key step in the global carbon cycle that regulates the atmosphere's composition while producing the food we eat and the fuels we burn. Approximately one-third of global carbon fixation occurs in an overlooked algal organelle called the pyrenoid. The pyrenoid contains the CO_2-fixing enzyme Rubisco, and enhances carbon fixation by supplying Rubisco with a high concentration of CO_2. Since the discovery of the pyrenoid over 130 years ago, the molecular structure and biogenesis of this ecologically fundamental organelle have remained enigmatic. Here, we use the model green alga *Chlamydomonas reinhardtii* to discover that a low complexity repeat protein, Essential Pyrenoid Component 1 (EPYC1), links Rubisco to form the pyrenoid. We find that EPYC1 is of comparable abundance to Rubisco and colocalizes with Rubisco throughout the pyrenoid. We show that EPYC1 is essential for normal pyrenoid size, number, morphology, Rubisco content and efficient carbon fixation. We explain the central role of EPYC1 in pyrenoid biogenesis by finding that EPYC1 binds Rubisco to form the pyrenoid matrix. We propose two models where EPYC1's four repeats could produce the observed lattice arrangement of Rubisco in the *Chlamydomonas* pyrenoid. Our results suggest a surprisingly simple molecular mechanism for how Rubisco can be packaged to form the pyrenoid matrix, potentially explaining how Rubisco packaging into a pyrenoid could have evolved across a broad range of photosynthetic eukaryotes through convergent evolution. Additionally, our findings represent a key step towards engineering a pyrenoid into crops to enhance their carbon fixation efficiency.

Significance statement

Eukaryotic algae, which perform approximately 40% of global net CO_2 fixation, enhance the performance of the carbon-fixing enzyme Rubisco by placing it into an organelle called the pyrenoid. Despite the ubiquitous presence and biogeochemical importance of this organelle, how Rubisco assembles to form the pyrenoid remains a longstanding mystery. Our discovery of an abundant repeat protein that binds Rubisco in the pyrenoid represents a critical advance in our understanding of pyrenoid biogenesis. The repeat sequence of this protein suggests elegant models to explain the structural arrangement of Rubisco enzymes in the pyrenoid. Beyond advances in our basic understanding, our findings open doors to the engineering of algal pyrenoids into crops to enhance yields.

Rubisco, the most abundant enzyme in the biosphere (1), fixes CO_2 into organic carbon that supports nearly all life on Earth (2, 3). Over the past three billion years, the enzyme became a victim of its own success, as it drew down the atmospheric CO_2 concentration to trace levels (4) and as the oxygen-producing reactions of photosynthesis filled our atmosphere with O_2 (4). In today's atmosphere, O_2 competes with CO_2 at Rubisco's catalytic site, producing toxic side-products approximately once every three catalytic cycles (5). These undesired side-products must be metabolized at the expense of energy and loss of fixed carbon and nitrogen (6). To overcome Rubisco's limitations, many photosynthetic organisms have evolved carbon-concentrating mechanisms (CCMs) (7, 8). CCMs increase the CO_2 concentration around Rubisco, decreasing O_2 competition and enhancing carbon fixation.

Approximately 28-44% of global carbon fixation is mediated by a CCM built around a poorly characterized organelle called the pyrenoid (3, 9-13), found in most oceanic eukaryotic algae (Table S1). The pyrenoid is a spherical structure in the chloroplast stroma, which for the past 25 years has been thought to primarily contain Rubisco and its chaperone, Rubisco activase (14). The molecular structure and biogenesis of the pyrenoid have remained enigmatic since its discovery over 130 years ago (15-17).

Results

EPYC1 is an abundant pyrenoid component. We hypothesized that the pyrenoid contains unidentified components that are important for its biogenesis. We therefore used mass spectrometry to analyze the protein composition of the pyrenoid of the model green alga *Chlamydomonas reinhardtii*, before and after applying a stimulus that induces pyrenoid growth. When cells are transferred from high CO_2 (2-5% CO_2 in air) to low CO_2 (air; ~0.04% CO_2), the CCM is induced (18) and the pyrenoid increases in size (19). We developed a protocol for isolating largely intact pyrenoids by cell lysis and centrifugation, and applied this protocol to cells before and after a shift from high to low CO_2 (Fig. 1*A*, Fig. S1*A*-*C*). Mass spectrometry indicated that the most abundant proteins in the low CO_2 pyrenoid fraction included the Rubisco large (rbcL) and small (RBCS) subunits, as well as Rubisco activase (RCA1; Fig. 1*B*, Fig. S1*D* and Dataset S1).

Strikingly, a fourth protein, previously identified as a low CO_2 induced protein (LCI5, Cre10.g436550) (20), was found in the low CO_2 pyrenoid fraction with comparable abundance to Rubisco (Fig. 1*B*). Based on the data presented herein, we propose to name this protein Essential Pyrenoid Component 1 (EPYC1). Under low CO_2, the stoichiometry of EPYC1 was ~1:6 with rbcL and ~1:1 with RBCS (estimated by intensity-based absolute quantification, iBAQ (21)). Consistent with EPYC1 being a component of the pyrenoid, the abundance of EPYC1 in the pyrenoid fraction increased ~12-fold after the shift from high to low CO_2 (Fig. 1*B*, Fig. S1*D* and Dataset S1), a similar increase to that of rbcL (7-fold), RBCS (7-fold), and RCA1 (19-fold). To confirm the pyrenoid localization of EPYC1, we fluorescently

tagged both EPYC1 and RBCS. Venus-tagged EPYC1 showed clear co-localization with mCherry-tagged RBCS in the pyrenoid (Fig. 1*C* and Fig. S1*E*).

EPYC1 is essential for a functional CCM. The high abundance of EPYC1 in the pyrenoid led us to ask whether EPYC1 is required for the CCM. We isolated a mutant in the 5' UTR of the *EPYC1* gene (Fig. S2*A* and Table S2), which contains markedly reduced levels of *EPYC1* mRNA (Fig. S2*B* and Table S3) and EPYC1 protein (Fig. 2*A*), and lacks transcriptional regulation in response to CO_2 (Fig. S2*B*). Similarly to previously described mutants in other components of the CCM, the *epyc1* mutant showed defective photoautotrophic growth in low CO_2, which was rescued by high CO_2 and by re-introducing the *EPYC1* gene (Fig. 2*B* and Fig. S2*C-E*). We further tested the CCM activity in *epyc1* mutants by measuring whole-cell affinity for inorganic carbon, inferred from photosynthetic O_2 evolution. When grown under low CO_2, the *epyc1* mutant showed a reduced affinity for inorganic carbon (increased $K_{0.5}$) relative to wild-type (WT; P=0.0055, n=5, Student's *t*-test; Fig. 2*C*, Fig. S2*F* and Table S4). These data indicate that EPYC1 is an essential component of the CCM.

EPYC1 is required for normal pyrenoid size and matrix density. Knowing that EPYC1 is in the pyrenoid and is required for the CCM, we explored whether the *epyc1* mutant shows any visible defects in pyrenoid structure. Thin-section transmission electron micrographs (TEM) revealed that *epyc1* mutants had smaller pyrenoids than WT at both low and high CO_2 (low CO_2: n=37-79, $P<10^{-19}$, Welch's *t*-test; high CO_2: n=18-22, $P<10^{-5}$, Welch's *t*-test; Fig. 3 *A* and *B*, Fig. S3 *A* and *B*, and Fig. S4). *Chlamydomonas* typically has one pyrenoid per cell (13). The *epyc1* mutant showed a higher frequency of multiple pyrenoids: 13% of non-dividing *epyc1* cells (n=231) showed multiple pyrenoids, compared with 3% of WT cells (n=252, P=0.00048, Fisher's exact test of independence, Table S5). Higher resolution quick-freeze deep-etch electron microscopy indicated a lower packing density of granular material in the pyrenoid matrix of the *epyc1* mutant (Fig. 3*C*, Fig. S3*C* and Fig. S5). This defect was most noticeable when cells were grown in low CO_2, but was also visible at high CO_2. Interestingly, the *epyc1* mutant retains a number of canonical pyrenoid characteristics (13), including correct localization, the presence of a starch sheath, and traversing membrane tubules, suggesting that these characteristics are regulated by mechanisms other than EPYC1.

EPYC1 is required for Rubisco assembly into the pyrenoid. Our observations of decreased pyrenoid size and apparent matrix density in *epyc1* mutants could be explained by decreased whole-cell levels of Rubisco. However, western blotting revealed no detectable difference in rbcL and RBCS abundance in *epyc1* relative to WT cells or between cells grown at low and high CO_2 levels (Fig. 3*D* and Fig. S3*D*). This result led us to hypothesize that the localization of Rubisco was perturbed in *epyc1* mutants. To test

this hypothesis, we generated WT and *epyc1* cell lines expressing Rubisco tagged with mCherry, and determined the distribution of fluorescence signal by microscopy. Remarkably, a large fraction of Rubisco was found outside the pyrenoid in the *epyc1* mutant. In *epyc1* cells grown in low CO_2, 68% of fluorescence from Rubisco tagged with mCherry was found outside the pyrenoid region, compared with 21% in WT cells (n=27, $P<10^{-15}$, Student's *t*-test; Fig. 3 *E* and *F* and Fig. S6). Immunogold-EM confirmed the mislocalization of Rubisco in *epyc1*. In pyrenoid-containing sections of low-CO_2-grown *epyc1* cells, 42% of anti-Rubisco immunogold particles were found outside the pyrenoid, whereas only 6% were found outside the pyrenoid in WT (WT: n= 26 cells, 8123 gold particles; *epyc1*: n=27 cells, 2708 gold particles; $P<10^{-15}$, Student's *t*-test; Fig. 3 *G* and *H* and Fig. S7).

If EPYC1 functions in the recruitment of Rubisco to the pyrenoid solely at low CO_2 (19), the *epyc1* mutant could be trapped in a "high-CO_2" state of Rubisco localization (19). However, the *epyc1* mutant showed a defect in Rubisco localization even under high CO_2 (Fig. S3 *E* and *F* and Fig. S6), where the fraction of Rubisco-mCherry fluorescence outside the pyrenoid region increased to 80% in *epyc1*, compared with 68% in the WT (WT: n=20, *epyc1*: n=20, $P=10^{-6}$, Student's *t*-test). We conclude that EPYC1 is required for Rubisco localization to the pyrenoid not only at low CO_2, but also at high CO_2.

EPYC1 and Rubisco are part of the same complex. EPYC1 could promote Rubisco's localization to the pyrenoid by a physical interaction. We therefore immunoprecipitated EPYC1 and Rubisco, and probed the eluates by western blotting (Fig. 4*A* and Fig. S8*A*). Immunoprecipitation of tagged EPYC1 pulled down the Rubisco holoenzyme; and reciprocally, tagged RBCS1 pulled down EPYC1. We conclude that EPYC1 and Rubisco are part of the same supramolecular complex in the pyrenoid. The high abundance of EPYC1 in the pyrenoid, its physical interaction with Rubisco, and its recruitment of Rubisco to the pyrenoid, all suggest that EPYC1 plays a structural role in pyrenoid biogenesis.

The EPYC1 protein consists of 4 nearly identical repeats. To gain insights into how EPYC1 might mediate the formation of such a network, we performed a detailed analysis of the EPYC1 protein sequence. This analysis indicated that EPYC1 consists of four nearly identical ~60 amino acid repeats (Fig. 4*B-D*), flanked by short N- and C-termini (in contrast to a previous study suggesting only three repeats (22)). We found that each repeat consists of a predicted disordered domain and a shorter, less disordered domain containing a predicted alpha helix (Fig. 4*C* and Fig. S8 *B* and *C*). Given that these repeats cover >80% of the EPYC1 protein, it is likely that the Rubisco binding site(s) are contained within the repeats.

Proteins with similar physicochemical properties to EPYC1 are present in a diverse range of eukaryotic algae. The primary sequences of disordered proteins like EPYC1 are known to evolve rapidly

compared to structured proteins, but their physicochemical properties are under selective pressure and are evolutionarily maintained (23). We therefore searched for proteins with similar physicochemical properties (repeat number, length, high isoelectric point, and disorder profile) across a broad range of algae (Table S6). Excitingly, proteins with similar properties are found in pyrenoid-containing algae, and appear to be absent from pyrenoid-less algae, suggesting that EPYC1-like proteins may play similar roles in pyrenoids across eukaryotic algae.

We propose two models for Rubisco assembly into the pyrenoid matrix by EPYC1. If each repeat of EPYC1 binds Rubisco, EPYC1 could link multiple Rubisco holoenzymes together into a hexagonal close packed or cubic close packed arrangement that could expand indefinitely in all directions, consistent with recent cryo-electron tomography studies of the *Chlamydomonas* pyrenoid (24). EPYC1 and Rubisco could interact in one of two fundamental ways: 1) EPYC1 and Rubisco could form a co-dependent network (Fig. 4*E*); or 2) EPYC1 could form a scaffold onto which Rubisco binds (Fig. 4*F*). Importantly, the 60 amino acid repeat length of EPYC1 is sufficient to span the observed 2-4.5nm gap between Rubisco holoenzymes in the pyrenoid (24), and a stretched out repeat could potentially span the 15nm observed Rubisco center-to-center distance. A promising candidate for an EPYC1 binding site on Rubisco would be the two alpha-helices of the small Rubisco subunit. When these helices are exchanged for higher-plant alpha-helices, pyrenoids fail to form and the CCM does not function, but holoenzyme assembly and *in vitro* kinetics are normal (25).

Discussion

The observations presented here suggest that Rubisco packaging to form the matrix of the eukaryotic pyrenoid is achieved by a different mechanism from that used in the well-characterized prokaryotic carbon-fixing organelle, the β-carboxysome. In the β-carboxysome, aggregation of Rubisco is mediated by the protein CcmM. CcmM contains multiple repeats of a domain resembling the Rubisco small subunit, and incorporation of these domains into separate Rubisco holoenzymes is thought to produce a link between Rubisco holoenzymes (26). Given that the EPYC1 repeats show no homology to Rubisco and are highly disordered, it is likely that they bind to the surface of Rubisco holoenzymes rather than becoming incorporated in the place of small subunits. The simplicity of such a surface binding mechanism potentially explains how Rubisco packaging into a pyrenoid could have evolved across a broad range of photosynthetic eukaryotes through convergent evolution (13, 27), leading to the dominant role of pyrenoids in aquatic CO_2 fixation.

In addition to being a key structural component, EPYC1 could be a central regulator of pyrenoid biogenesis. The Rubisco content of the pyrenoid changes in response to light (28) and CO_2 ((19) and our data). This change in Rubisco localization could be mediated by changes in EPYC1 abundance and/or

Rubisco binding affinity. EPYC1 was previously found to be upregulated at both the transcript and protein levels in response to light and low CO_2 (22), and our data further supports this finding (Fig. 2*A* and Fig. S2*A*). Moreover, EPYC1 becomes phosphorylated at multiple sites in response to low CO_2 (22, 29), potentially affecting its binding affinity for Rubisco. Several of the phosphorylation sites include serine-proline and threonine-proline motifs, which are known to induce conformational changes (30). Phosphorylation of these motifs could regulate the structure of EPYC1, potentially affecting EPYC1-Rubisco binding or the distance between linked Rubisco holoenzymes. Another mode of regulation could be by methylation of Rubisco or EPYC1, as the predicted methyltransferase CIA6 is required for Rubisco localization to the pyrenoid (31).

Further to advancing our understanding of the molecular mechanisms underlying global carbon fixation, our findings may aid to enable the engineering of crops with enhanced photosynthesis. There is great interest in introducing a CCM into C_3 plants, as this enhancement is predicted to increase yields by up to 60% and improve nitrogen and water use efficiency (32). Our discovery of a possible mechanism for pyrenoid formation is a key step towards engineering a pyrenoid, a central component of the CCM, into crops.

Methods

Strains and culture conditions

Wild-type (WT) *Chlamydomonas reinhardtii* CC-1690 (33) was used for pyrenoid enrichment and proteomics. This strain was maintained at 22°C with 55 µmol photons m^{-2} s^{-1} light on TAP (34) agar (1.4%) plates containing 0.4% Bacto-Yeast extract (Becton, USA). For the localization of EPYC1 in the absence of Rubisco, strain CC-4415 (*rbcS*-T60-3) which lacks both the Rubisco small subunit genes (35) was used. For all other experiments, *Chlamydomonas* WT strain cMJ030 (CC-4533) (36) was used. The background strain of the *epyc1* mutant is cMJ030. The cMJ030 WT and *epyc1* mutant were maintained in the dark or low light (~10 µmol photons m^{-2} s^{-1}) on 1.5% agar plates containing tris-acetate-phosphate (TAP) with revised (37) or traditional Hutner's trace elements (38).

For proteomics analysis, a 50 mL pre-culture was grown mixotrophically in TAP on a rotatory shaker at 124 rpm, 22°C and under an illumination of 55 µmol photons m^{-2} s^{-1} for three days according to Mettler*, et al.* (39). In brief, a second pre-culture of 500 mL was used to inoculate a 5-litre bioreactor BIOSTAT®B-DCU (Sartorius Stedim, Germany). The absence of contamination was monitored according to Mettler*, et al.* (39). Cultures with a cell density of 3-5 x 10^6 cells mL^{-1} were grown photoautotrophically at 46 µmol photons m^{-2} s^{-1} light in air enriched with high CO_2 (5% CO_2) under constant turbidity for two days before the culture was aerated with low CO_2 (ambient air; 0.039% CO_2). The CO_2 level in the outlet air of the bioreactor was measured by an on-line multi-valve gas chromatograph (3000A MicroGC run by EZChromElute software, Agilent Technologies, USA). After switching from high to low CO_2, the CO_2 dropped from 4.5% to a constant 0.02% after 12 hours. Cells were harvested at 30 hours after the shift to low CO_2.

For mRNA levels, O_2 evolution, Rubisco content western blotting, pyrenoid size analysis by transmission electron microscopy (TEM) and Rubisco subcellular localization by immuno-gold labelling, strains were grown photoautotrophically in 50 mL tris-minimal medium (38) under constant aeration, shaking and illumination (150 rpm, 21°C, 50-65 µmols photons m^{-2} s^{-1}; Infors HT Multitron Pro, Switzerland). Briefly, starter cultures were inoculated from freshly re-plated cultures on TAP plates, to 0.3 µg chlorophyll (a+b) mL^{-1}, and aerated with high CO_2 (5% v/v CO_2 enriched air). When cell density reached mid-log (~3 µg chlorophyll (a+b) mL^{-1}), half of the cultures were harvested and analysed. The remaining half of the cultures were then switched to aeration with low CO_2 (0.04% v/v CO_2) for induction of the CCM. For gene expression analysis and affinity for inorganic carbon, cells were air-adapted for 3 hours, corresponding to peak induction of CO_2-inducible genes (28, 40, 41). The state of CCM induction was controlled by measuring the mRNA accumulation of a highly CO_2-responsive gene, *LCI1*

(Cre03.g162800). For TEM analysis of pyrenoid size and immuno-gold labelling of Rubisco, cells were adapted to low CO_2 for 24 hours.

For EPYC1 protein abundance and freeze fracture cryo-electron microscopy of WT and mutant cells, cultures were propagated continuously in tris-phospate (TP) (37) medium with 50 µmols photons m^{-2} s^{-1} light for ~1 week in a Multi-Cultivator (Photon Systems Instruments) with bubbling of high CO_2 (3% v/v CO_2). Cells were diluted every 24 hours to ensure they were kept in the log phase. 6 hours before sampling, half of the cultures were switched from high CO_2 to low CO_2 (~0.04% v/v CO_2). The chlorophyll concentration at harvesting was ~3 µg chlorophyll (a+b) mL^{-1}.

For fluorescence microscopy and RbcS1-mCherry localization experiments, cells were grown in TP medium containing antibiotics used for selection of expression of the fluorescently labeled gene (Venus, paromomycin at 2 µg mL^{-1}; mCherry, hygromycin 6.25 µg mL^{-1}), bubbled with high CO_2 (3% v/v CO_2) at a 150 µmol photons m^{-2} s^{-1} light intensity. At ~2 x 10^6 cells mL^{-1}, after >6 doublings, cells were transferred to low CO_2 for 14 hours. For the RbcS1-mCherry localization experiments, samples were taken and imaged immediately before the switch to low CO_2 and after 14 hours at low CO_2.

For co-immunoprecipitation experiments, cells were grown in 50 mL of TAP at 150 µmol photons m^{-2} s^{-1} light until ~2-4 x 10^6 cells mL^{-1}, centrifuged at 1000 *g* for 4 min, resuspended in TP and used to inoculate 800 mL of TP. Cells were then bubbled with low CO_2 (air, ~0.04% v/v CO_2) at 150 µmol photons m^{-2} s^{-1} until ~2-4 x 10^6 cells mL^{-1} and harvested as indicated below. All liquid media contained 2 µg mL^{-1} paromomycin.

Cell concentrations were measured using a Z2 Coulter Counter (Beckman Coulter, USA).

Proteomics of pyrenoid-enriched fraction

Transmission electron microscopy

TEM images of whole cells for pyrenoid enrichment and the enriched pyrenoid fraction were prepared and taken according to Nordhues*, et al.* (42).

Pyrenoid enrichment

10 mL algal material (3-5 x 10^6 cells mL^{-1}) were harvested by centrifugation for 2 min (4,000 rpm, 4°C), immediately frozen in liquid nitrogen and extracted with extraction buffer (EB; 50 mM HEPES, 20µM leupeptin, 1 mM PMSF, 17.4% glycerol, 2% Triton). The samples were sonicated 6 x 15 s (6 cycles, 50% intensity, Sonoplus Bandelin Electronics, Germany) and kept on ice between cycles for 90 s. The samples were centrifuged at 500 *g* for 3 min to obtain a soluble and pellet fraction. This procedure resembled the first steps of a protocol used in previous studies (43, 44). The pellet was washed three times with 1 mL,

500 µL and 300 µL EB before resuspension in 100µL 50 mM ammonium bicarbonate. Protein concentrations were measured by Lowry assay using BSA as a standard (45).

SDS-PAGE

For SDS-PAGE, samples were resuspended in a buffer containing 50 mM dithiothreitol (DTT), 50 mM sodium-carbonate, 15% sucrose (w/v) and 2.5% SDS (w/v), heated 45 seconds at 95°C and spun down at 14,000 rpm before applying 22 µg total protein to the polyacrylamide gel. The 14%-separating gel was stained with Coomassie Brilliant Blue (46).

Protein digestion and mass spectrometric analysis

For shotgun proteomics, samples were prepared and measured according to Mühlhaus *et al.* (47). In brief, 20 µg protein per sample was precipitated in 80% acetone at -20°C over night. The precipitated proteins were resuspended in 6 M urea and 2 M thiourea (in 50 mM ammonium hydrogen carbonate), reduced by DTT, carbamidomethylated with iodoacetamide, digested with endoproteinase LysC (Roche, Switzerland) and immobilised trypsin (Applied Biosystems, Thermo Fisher Scientific, USA), and subsequently desalted according to Mühlhaus *et al.* (47). The resuspended peptides were acidified with 1% acetic acid. Peptides were chromatographically separated by reverse phase separation with a nanoUPLC (nanoACQUITY UPLC, Waters, Milford, USA) using a 10cm x 75µm BEH130 C18 1.7µm particles (Waters) column for separation and a 2cm x 180µm Symmetry C18 5µm particles (Waters) column for trapping. Peptides were analysed by a linear trap quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, USA) according to Mühlhaus *et al.* (47).

Data processing and data analysis

Raw MS files were processed with MaxQuant (version 1.5.2.8) (48). Peak list files were searched against *Chlamydomonas reinhardtii* gene model JGIv4 from Phytozome 10.2 (http://phytozome.jgi.doe.gov/) including the organelle genome sequences. Maximum precursor and fragment mass deviations were set to 20 ppm and 0.5 Da. Peptides with at least six amino acids were considered for identification. The search included carbamidomethylation as a fixed modification and variable modifications for oxidation of methionine and protein N-terminal acetylation. The false discovery rate, determined by searching a reverse database, was set at 0.01 for both peptides and proteins. Identification across different replicates and treatments was achieved by enabling the "match between runs" option in MaxQuant within a time window of 2 min. For comparison of protein levels between samples, the label-free quantification (LFQ) intensity based method was used (48). For the estimation of protein stoichiometries within a sample, the intensity-based absolute quantification (iBAQ) (21) method was applied. Both values were calculated by the MaxQuant software. All statistical analyses were performed using Microsoft Excel.

Cloning of EPYC1 and RbcS1

EPYC1 (Cre10.g436550) and *RBCS1* (Cre02.g120100) ORFs were amplified from gDNA using Phusion Hotstart II polymerase (Thermo Scientific) with the respective EPYC1_ORF_F/R or RBCS1_ORF_F/R primer pairs (Table S2). Gel purified PCR products, containing vector overlap regions, were cloned into pLM005 or pLM006 by Gibson assembly (49). Final pLM005 constructs are in frame with Venus-3xFLAG and contain the AphVIII gene for paromomycin resistance, final pLM006 constructs are in frame with mCherry-6xHIS and contain the AphVII gene for hygromycin resistance. Both pLM005 and pLM006 confer ampicillin resistance for bacterial selection. For complementation with untagged EPYC1, mCherry-6XHIS was removed from pLM006_EPYC1-mCherry-6xHIS by BglII restriction digestion, gel purified then re-ligated. All constructs were verified by Sanger sequencing.

Transformation of *Chlamydomonas* for complementation and fluorescence localization of proteins

WT and *epyc1* strains were transformed by electroporation as in Zhang*, et al.* (36). For each transformation, 14.5 ng kbp^{-1} of EcoRV cut plasmid was mixed with 250 µL of 2 x 10^8 cells mL^{-1} at 16 °C and transformed immediately. Cells we selected on TAP paromomycin (20 µg mL^{-1}) or hygromycin (25 µg mL^{-1}) plates and kept in low light (5-10 µmol photons m^{-2} s^{-1}) until picking or screening for fluorescence lines. In addition, for the complementation of the *epyc1* mutant, a second tranformation was selected on TP plates, without antibiotics at low CO_2 (~0.04% v/v CO_2) under 500 µmol photons m^{-2} s^{-1} light. Transformation of CC-4415 was also performed as above, except all growth and recovery steps were in the dark.

To screen for Venus and mCherry expressing lines, transformations were spread on rectangular plates (Singer Instruments) containing 86 mL of TAP plus antibiotics. Once colonies were ~2-3 mm in diameter, plates were transferred to ~100 µmol photons m^{-2} s^{-1} light for 24-36 hours and then screened for colony fluorescence on a Typhoon TRIO fluorescence scanner (GE Healthcare). Excitation and emission settings were: Venus, 532 excitation with 555/20 emission; mCherry, 532 excitation with 610/30 emission; and chlorophyll autofluorescence, 633 excitation with 670/30 emission. Dual-tag lines were generated sequentially by expressing pLM005_EPYC1-Venus-3xFLAG in WT then adding pLM006_RbcS1-mCherry-6xHIS. To confirm expression of both Venus and mCherry in dual-tag strains and to select for strains with equal fluorescence intensity for the analysis of RbcS1-mCherry localization in WT and *epyc1,* strains were also screened on a Tecan Infinite M1000 PRO (50).

Fluorescence microscopy and Rubisco-mCherry mislocalization in the *epyc1* mutant

All fluorescence microscopy was performed using a spinning disk confocal microscope (custom adapted Leica DMI6000) with samples imaged on poly-L-lysine coated plates. The following excitation and emission settings were used: Venus, 514 excitation with 543/22 emission; mCherry, 561 excitation with 590/20 emission; and chlorophyll, 561 excitation with 685/40 emission. Images were analysed using Fiji software. For RbcS1-mCherry localization in WT and the *epyc1* mutant, lines showing equal RbcS1-mCherry fluorescence intensity were chosen for analysis (see above). WT and *epyc1* lines were imaged using the above mCherry and chlorophyll settings. A Z-stack composed of 40 slices 0.3 µm apart was obtained for each field of view. To quantify the percentage of fluorescence signal from outside the pyrenoid region, raw images were analysed as follows: Pixel intensity in the mCherry channel was summed across the 40 Z-sections for cells that were fully sectioned. Using the chlorophyll channel as a reference a cell outline region of interest (ROI; varying between cells) and pyrenoid ROI (set at 2.8 µm in diameter for WT and mutant) were drawn. For each cell, background fluorescence was subtracted by taking the average of 4 measurements surrounding the cell, and autofluorescence was subtracted separately from the pyrenoid and whole cell ROIs by taking the average of 22 WT cells not expressing mCherry. Finally, the percentage RbcS1-mCherry signal from outside of the pyrenoid region was calculated as the (*total cell signal* - *pyrenoid signal*) / *total cell signal* x 100%.

Analysis of gene expression by qRT-PCR

Quantitative real-time PCR was used to determine the relative abundance of EPYC1 gene transcripts. Total RNA was extracted from 30 µg chlorophyll a+b (~10 mL mid-log cell suspension), using TRIzol Reagent, as per manufacturer's instructions (Life Technologies). Complementary DNA was synthesised from 500 ng of total RNA using SuperScript III Reverse Transcriptase (Life Technologies), RNaseOUT (Life Technologies), and oligo(dT)18 primers (Thermo Scientific). Relative gene expression was measured in real time in a Rotor-Gene Q thermocycler (Qiagen). Reactions (10 µL) used SYBR Green JumpStart Taq ReadyMix (Sigma-Aldrich). Gene expression was calculated according to the method of Livak and Schmittgen (51), relative to the *Chlamydomonas* gene coding for the Receptor of Activated Protein Kinase C1 (RCK1, Cre06.g278222) (52), which is not significantly induced by low-CO_2 (41). All primers used are in Table S2.

Screening for the *epyc1* mutant

The *epyc1* mutant was isolated from a collection of high CO_2 requiring mutants by a pooled screening approach. A collection of approximately 7,500 mutants on 79 plates, each with 96 colonies, was grown in

liquid TAP in 96 well plates then pooled by well row, well column, whole plate row and whole plate column to give a total of 38 pools. Pooled cells were pelleted, DNA was extracted by phenol:chloroform:isoamyl alcohol (Phenol:CIA, 25:24:1; Sigma-Aldrich) and then screened by PCR for an EPYC1 mutant using a primer in the pMJ016c mutagenesis cassette (a modified pMJ013c cassette) (36) and a primer in the EPYC1 gene. The identified *epyc1* mutant has an insertion of the pMJ016c resistance cassette in the 5'UTR, the resistance cassette is 11 bp upstream of the ATG start codon, with the cassette having a 10 bp deletion at the 3' end. The upstream gDNA-cassette junction cannot be PCR amplified. However, PCR shows the full cassette is intact and that >397 bp upstream of the insertion site is also intact (Fig S2*A*). All primers used are in Table S2.

Protein extraction and western blotting

For EPYC1 protein quantification in WT and the *epyc1* mutant, protein was extracted from unfrozen cells, normalised to chlorophyll, separated by SDS-PAGE and western blotted as described in Heinnickel*, et al.* (53). The primary anti-EPYC1 antibody was used at a 1:2,000 concentration and the secondary horseradish-peroxidase (HRP) conjugated goat anti-rabbit (Life Technologies) at a 1:10,000 concentration. To ensure even loading, membranes were stripped (Restore PLUS western blot stripping buffer, Thermo Scientific) and re-probed with anti-tubulin (1:25,000; Sigma) followed by HRP conjugated goat anti-mouse (1:10,000; Life Technologies). The anti-EPYC1 antibody was raised in rabbit to the C-terminal region of EPYC1 (KSKPEIKRTALPADWRKGL-cooh) by Yenzym Antibodies (South San Francisco, California, USA).

For Rubisco quantification in WT and *epyc1* mutant, total soluble proteins were extracted from 300 µg chlorophyll (a+b) (~100 mL mid-log cell suspension). Cells were harvested by centrifugation (13,000 *g*, 10 min, 4°C), re-suspended in ice cold 1.5 mL extraction buffer (50 mM Bicine, pH 8.0, 10 mM $NaHCO_3$, 10 mM $MgCl_2$, and 1 mM DTT), and lysed by sonication (6 x 30 second bursts of 20 microns amplitude, with 15 s on ice between bursts; Soniprep 150, MSE UK Ltd, London, UK,). Lysis was checked by inspecting samples under a light microscope. Lysate was clarified by centrifugation (13,000 *g*, 20 min, 4°C). Protein content was determined using the Bradford method (Sigma Aldrich). Soluble proteins were separated on 12% (w/v) denaturing polyacrylamide gel. Sample loading was normalised by protein amount (10 µg per lane), and even loading was controlled by staining a gel with identical protein load (GelCode Blue, Life Technologies). After transfer onto a polyvinylidene difluoride membrane (Amersham), Rubisco was immuno-detected with a polyclonal primary antibody raised against Rubisco (1:10,000) followed by a HRP conjugated goat anti-rabbit (1:20,000; GE Healthcare).

Chlorophyll concentration

Total pigments were extracted in 100% methanol, and the absorbance of the clarified supernatant (13,000 *g*, 1 min, 4°C) was measured at 470, 652, 665, 750 nm (UV 300 Unicam, Thermo Spectronic, Cambridge, UK). Concentration of chlorophyll (a+b) was calculated using the equation of Wellburn (54).

Spot tests

WT, *epyc1* and complemented cell lines were grown in TAP until ~2x10^6 cells mL^{-1}, washed once with TP, resuspended in TP to a concentration of 6.6x10^5 cells mL^{-1}, then serially diluted 1:10 three times. 15 µL of each dilution was spotted onto four TP plates and incubated in low or high CO_2 with 100 or 500 µmol photons m^{-2} s^{-1} of light for seven days before imaging.

Oxygen evolution measurements

Apparent affinity for inorganic carbon was determined using the oxygen evolution method described by Badger, Kaplan and Berry (55). Photoautotrophically grown liquid cultures were harvested by centrifugation (2,000 *g*, 5 min, 4°C) and re-suspended in 25 mM HEPES-KOH (pH 7.3) to a density of ~1.5 x 10^8 cells mL^{-1}, as determined by haemocytometer count. Aliquots of cells (1 mL) were added to a Clark-type oxygen electrode chamber (Rank Brothers, Bottisham, UK) attached to a circulating water bath set to 25°C. The chamber was closed for a light pre-treatment (200-300 µmol photons m^{-2} s^{-1} illumination for 10-25 min), to allow cells to deplete any internal inorganic carbon pool. When net oxygen evolution ceased, 10 µL of increasingly concentrated $NaHCO_3$ solution was added to the algal suspension at 30 second intervals, and the rate of oxygen evolution was recorded every second using a PicoLog 1216 data logger (Pico Technologies, St Neots, UK). Cumulative concentrations of sodium bicarbonate after each addition were as follows: 2.5, 5, 10, 25, 50, 100, 250, 500, 1,000, and 2,000 mM. Michaelis-Menten curves were fitted to plots of external inorganic carbon concentration versus the rate of O_2 evolution. The concentration of inorganic carbon required for half maximal rates of photosynthesis ($K_{0.5}$) was calculated from this curve.

Pyrenoid area analysis by transmission electron microscopy

To minimise the loss of biological signal during harvesting, fixative (glutaraldehyde, final 2.5%) was added to cell cultures immediately before harvesting. Cell suspensions containing ~5 x 10^7 cells in mid-log were pelleted (4,000 *g*, 5 min, 4°C) and fixed in 1 mL tris-minimal medium containing 2.5% glutaraldehyde and 1% H_2O_2 (30% w/v) for 1 hour on a tube rotator at 4°C. Unless otherwise specified,

all following steps were performed at room temperature on a tube rotator. Cells were pelleted (4,000 *g* 5 min) and washed with ddH_2O (3X, 5 min). Cells were osmicated for 1 hour in 1 mL 1% (v/v) OsO_4 containing 1.5% (w/v) $K_3[Fe(CN)_6]$ and 2 mM $CaCl_2$. Cells were pelleted and washed with ddH_2O (4X, as above). Cells were stained for 1 hour in 1 mL 2% (w/v) uranyl acetate. After pelleting and washing with ddH_2O (3X), cells were dehydrated in 70%, 95%, 100% ethanol, and 100% acetonitrile (2X). Cells were embedded in epoxy resin mix, containing Quetol 651, nonenyl succinic anhydride, methyl-5-norbornene-2,3-dicarboxylic anhydride, and dimethyl-benzylamine (all reagents from Agar Scientific, Stansted, UK), in the following proportions: 35%, 46%, 17%, 2%. Resin was refreshed 4X over the two days. Thin sections (50 nm) were prepared by the Cambridge Advanced Imaging Centre (Ms Lyn Carter) on a Leica Ultracut UCT Ultramicrotome and mounted onto 300 mesh copper grids. Samples were imaged with a Tecnai G2 transmission electron microscope (FEI, Hilsboro, OR, USA) at 200 kV. Image analysis (area measurements) was performed using ImageJ. Ten 54 μm^2 areas were randomly selected and all pyrenoid positive cells were imaged (WT low CO_2, 79 out of 271 cells displayed a pyrenoid; *epyc1* low CO_2, 37 out of 139 cells displayed a pyrenoid; WT high CO_2, 18 out of 196 cells displayed a pyrenoid; *epyc1* high CO_2, 22 out of 255 cells displayed a pyrenoid). Cell area was determined by outlining the plasma membrane. Pyrenoid area was taken as the area inside the starch sheath (generally visible in CCM-induced cells) or the electron dense area inside the chloroplast when no starch sheath was visible. Control immuno-gold labelling experiments using a high concentration of primary antibody (1:20) confirmed that these areas had dense concentrations of Rubisco. Pyrenoid area was expressed as a percentage of cell area, and data was ordained in classes of 0.5% increment.

Quick-freeze deep-etch EM (QFDEEM)

Sampling and fixation

It was ascertained in pilot experiments that pyrenoids fixed by the following procedure are indistinguishable in QFDEEM ultrastructure from unfixed controls. 150 mL of each of air-bubbled cultures and 75 mL of high CO_2-bubbled cultures were pelleted at 1,000 *g* for 10 min at RT to produce pellets of ~200 µL. The pellets were resuspended in 6 mL of ice-cold 10 mM HEPES buffer (pH 7) and transferred to a cold 25 mL glass flask. A freshly prepared solution of 4% glutaraldehyde (Sigma-Aldrich G7651) in 10 mM HEPES (pH 7) was added 100 µL at a time, swirling between drops, until 1.5 mL in total had been added. The mixture was then left on ice for 1 hour, with agitation every 10 min. The mixture was pelleted (1000 *g*, 5 min, 4° C), washed in cold HEPES buffer, pelleted again, and finally resuspended in 6 mL fresh HEPES. Samples were shipped overnight to St. Louis in 15 mL conical screw cap tubes maintained at 0-4° C.

Microscopy

QFDEEM was performed as in Heuser (56). Briefly, small samples of pelleted cells were placed on a cushioning material and dropped onto a liquid helium-cooled copper block; the frozen material was transferred to liquid nitrogen and then to an evacuated Balzers apparatus, fractured, etched at -80°C for 2 min, and platinum/carbon rotary-replicated. The replicas were examined with a JEOL electron microscope, model JEM 1400, equipped with an AMTV601 digital camera. The images are photographic negatives; hence, protuberant elements of the fractured/etched surface are more heavily coated with platinum and appear whiter.

Immunogold-localization of Rubisco

Resin embedded material previously used for ultra-structural characterization of the pyrenoid was re-cut and thin sections were mounted on nickel grids. Superficial osmium and unmasking of epitopes was done by acid treatment (57). Grids were gently floated face down on a droplet (~30 µL) of 4% sodium meta-periodate (w/v in ddH_2O) for 15 min, and 1% periodic acid (w/v in ddH_2O) for 5 min. Each acid treatment was followed by several short washes in ddH_2O. Non-specific sites were blocked for 5 min in 1% BSA (w/v) dissolved in high-salt tris-buffered saline containing 500 mM NaCl, 0.05% Triton X-100 and 0.05% Tween 20 (hereafter abbreviated HSTBSTT). Salt, detergent, and surfactant concentrations were determined empirically to minimise background signal. Binding to primary antibody was done by incubating grids overnight in 1% BSA in HSTBSTT, with 1:1,000 dilution of the Rubisco antibody. Excess antibody was removed by 15 min washes (2X) in HSTBSTT and 15 min washes (2X) in ddH_2O. Incubation with secondary antibody (15 nm gold particle-conjugated goat anti-rabbit secondary antibody in 1% BSA in HSTBSTT, 1:250) was done at RT for 1 hr. Excess secondary antibody was removed by washing as above. Thin sections were prepared and imaged as for *Pyrenoid area analysis by transmission electron microscopy*, above. Randomisation was done as above (see TEM) with scoring capped to ~25 cells for each treatment. Non-specific labelling was taken as any particle on a free resin area, i.e. outside a cell. Non-specific density was subtracted from pyrenoid and chloroplast particle density. Fraction of particles in the pyrenoid was calculated as background-adjusted $n_{pyrenoid} / (n_{pyrenoid} + n_{stroma})$, where n_{stroma} is the number of particles in the stroma to the exclusion of the pyrenoid and the starch sheath. To improve the clarity of gold particles in Fig. 3g, particles were enlarged 10x using the image analysis software, Fiji. Briefly, images were thresholded to isolate individual gold particles, these were then enlarged 10x, and the new image overlaid on the original image with an opacity of 50%.

Co-Immunoprecipitations

WT cells expressing pLM005_Venus-3xFLAG, pLM005_EPYC1-Venus-3xFLAG or pLM005_RbcS1-Venus-3xFLAG were grown in 800 mL of TP plus 2 µg mL^{-1} paromomycin with continual bubbling at low CO_2 (0.04% CO_2) under 150 µmol photons m^{-2} s^{-1} of light until a cell density of ~2-4 x 10^6 cells mL^{-1}. Cells were then spun out (2,000 *g*, 4 min, 4°C), washed in 40 mL of ice cold TP, centrifuged then resuspended in a 1:1 (v/w) ratio of ice-cold 2xIP buffer (400 mM sorbitol, 100 mM HEPES, 100 mM KOAc, 4 mM $Mg(OAc)_2.4H_2O$, 2 mM $CaCl_2$, 2 mM NaF, 0.6 mM Na_3VO_4 and 1 Roche cOmplete EDTA-free protease inhibitor/ 25 mL) to cell pellet. This cell slurry was then added drop wise to liquid nitrogen to form small *Chlamydomonas* "popcorn" balls approximately 5 mm in diameter. These were stored at -70°C until needed.

Cells were lysed by grinding 1g (~500 mg of original cell pellet) of *Chlamydomonas* popcorn balls by mortar and pestle at liquid nitrogen temperatures, for 10 min. The ground cells were defrosted on ice, then dounced 20 times on ice with a Kontes Glass Co. Duall #21 homogeniser. Membranes were solubilised by incrementally adding an equal volume of ice-cold 1xIP buffer plus 2% digitonin (final concentration is 1%), then incubating at 4°C for 40 min with nutation. The lysate was then clarified by spinning for 30 min at full-speed in a table-top centrifuge at 4°C. The supernatant (Input) was then transferred to 225 µL of protein G Dynabeads (Life Technologies) that had been incubated with anti-FLAG M2 antibody (Sigma) according to the manufacturer's instructions, except 1xIP buffer was used for the wash steps. The Dynabead-cell lysate was incubated for 2.5 hours on a rotating platform at 4°C, then the supernatant removed (Flow-through). The Dynabeads were washed 4 times with 1xIP buffer plus 0.1% digitonin followed by a 30 min elution with 50 µL of 1xIP buffer plus 0.25% digitonin and 2 µg/ µL 3xFLAG peptide (Sigma; 3xFLAG peptide elution) and a 10 min elution in 1x Laemmli buffer with 50 mM beta-mercaptoethanol at 70°C (Boiling elution). Samples were run on 10% SDS-PAGE gels, then silver stained or transferred to PVDF membrane and probed with anti-FLAG (1:2,000; secondary: 1:10,000 HRP goat anti-mouse), anti-Rubisco (1:10,000; secondary: 1:20,000 HRP goat anti-rabbit) or anti-EPYC1 (1:2,000; secondary: 1:10,000 HRP goat anti-rabbit).

EPYC1 sequence analysis

To understand the intrinsic disorder of EPYC1, the full-length amino acid sequence was run through several structural disorder prediction programs including VL3, VLTX(58) and GlobPlot 2 (59). To look for regions of secondary structure, the full-length and repeat region of the EPYC1 amino acid sequence was analysed by PSIPRED v3.3 (60) and Phyre2 (61).

EPYC1-Rubisco interaction model

We built a model of the EPYC1-Rubisco interaction using Blender (www.blender.org) based on the following logic: If each of the 4 EPYC1 repeats can bind a holoenzyme, the 2 internal repeats would have different linking properties from the 2 terminal repeats. If bound to an internal repeat, a holoenzyme would be directly linked through this EPYC1 protein to 2 other holoenzymes. In contrast, if bound to a terminal repeat, the holoenzyme would only be directly linked through this EPYC1 protein to one other holoenzyme. Therefore on average, each EPYC1 repeat would link one Rubisco holoenzyme to 1.5 other holoenzymes. Given the octameric structure of the Rubisco holoenzyme, a holoenzyme likely has 8 binding sites for EPYC1. Taken together, on average each holoenzyme would be bound to 12 other holoenzymes by 8 EPYC1 proteins, in an arrangement that could expand indefinitely in all directions. A perfect arrangement of this nature would require a stoichiometry of one EPYC1 polypeptide for every four Rubisco small or large subunits.

Analysis of other algal proteomes for EPYC1-like physicochemical properties

Complete translated genomic sequences from pyrenoid and non-pyrenoid algae were downloaded from Uniprot or Phytozome. Protein sequences were then analysed for tandem repeats using Xstream (62) with default settings except: Min Period, 40; Max Period, 80; Min Copy #, 3.0; Min TR Domain, 75; Min Seq Content, 0.7. The pI of the Xstream hits were then batch calculated using the Gene Infinity Protein Isolelectric Point calculator (http://www.geneinfinity.org/sms/sms_proteiniep.html) and the disorder profile calculated using VLXT (63). Proteins with an oscillating disorder profile with a frequency between 40-80 were classified as potential Rubisco linker proteins.

Statistical methods

When growing algal material in liquid medium, flasks were placed randomly throughout the orbital shaker/incubator. Placement was randomized after each sub-culturing to offset any differences in illumination quality. Manifold for air/CO_2 delivery had standardized tubing length and internal diameter for even aeration. Cells lysis via sonication required samples to be processed sequentially. Order of processing was randomized.

Sample size of O_2 evolution measurement was aligned to previously published work from the Griffiths Lab (25, 28). Sample size of electron microscopy related experiments (scoring of TEM thin sections and immunogold experiments) was validated by jackknife resampling.

Pre-established exclusion criteria for TEM image scoring were: (i) only grid areas fully covered with material (54 μm^2) were considered; (ii) sections through broken cells and cell sections with a cross area < 12.5 μm^2 (a circle with 2 μm radius served as a guide) were not scored.

Scoring of electron micrographs: images files were renamed with a random number (RANDBETWEEN function in Microsoft Excel), sorted from high to low, and scored blindly. The original filename appearing on the bottom left of each micrograph was masked during the on-screen processing in ImageJ. Randomly selected imaged were scored by a second experimenter for independent validation. No systematic bias (over- or underestimation) was measured, and measurements deviated on average only by a couple of percentage points.

Two-tailed Student's *t*-test was used to compare affinities for inorganic carbon of WT and *epyc1*, as well as the mislocalization of Rubisco by fluorescence microscopy and EM, because this test is robust to non-normal distributions (64). Welch's *t*-test was used to compare pyrenoid sizes, because the WT and mutant groups had substantially different standard deviations (64). Fisher’s exact test of independence was used to compare the number of pyrenoids in WT and *epyc*, as this test is appropriate when there are two nominal variables (64).

Acknowledgments We thank J. Skepper, L. Carter and M. Rütgers for TEM support, discussions on immunogold optimisation, and ultramicrotomy; H. Cartwright for confocal microscopy support; S. Ramundo for technical advice with co-immunoprecipitations; W. Patena for help with data analysis; and W. Frommer, V. Walbot, P. Walter and T. Cuellar for comments on the manuscript. The project was funded by BBSRC grant BB/M007693/1 (M.T.M. and H.G.), NSF grants EF-1105617 and IOS-1359682 (L.C.M.M. and M.C.J.), the Carnegie Institution for Science (L.C.M.M. and M.C.J.), by the Federal Ministry of Education and Research (BMBF), Germany, within the frame of the *Go*FORSYS Research Unit for Systems Biology (FKZ 0313924; T.M.-A. and M.S.) and the International Max Planck Research School (IMPRS) of the Max Planck Society (T.M.-A.).

Author contributions L.C.M.M., M.T.M., T.M.-A., M.S., H.G. and M.C.J. designed experiments; T.M.-A. isolated pyrenoids; S.G. performed TEMs of pyrenoid-enriched fractions; T.M.-A., M. Schroda, F.S. and T. Mühlhaus analysed pyrenoids by mass spectrometry; L.C.M.M. determined the localization of fluorescently tagged proteins, performed mutant genotyping, complementation, spot tests, and immunoprecipitations; L.C.M.M., L.P. and G.R. isolated the *epyc1* mutant; L.C.M.M., V.C. and A.I. performed EPYC1 westerns; M.T.M. performed qRT-PCRs, TEMs, immunogold TEMs, and Rubisco westerns; M.C.M. measured O_2 evolution; E.S.F., V.C., R.R. and U.G. performed freeze-fracture EMs; V.C, L.C.M.M., M.T.M. and O.C. performed analyses of EPYC1 protein sequence. L.C.M.M., M.T.M., T.M.-A., H.G. and M.C.J. wrote the manuscript. All authors read and commented on the manuscript.

References

1. Ellis RJ (1979) The most abundant protein in the world. *Trends in biochemical sciences* 4(11):241-244.
2. Falkowski PG, Barber RT, & Smetacek V (1998) Biogeochemical controls and feedbacks on ocean primary production. *Science* 281(5374):200-206.
3. Field CB, Behrenfeld MJ, Randerson JT, & Falkowski P (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281(5374):237-240.
4. Dismukes G*, et al.* (2001) The origin of atmospheric oxygen on Earth: the innovation of oxygenic photosynthesis. *Proceedings of the National Academy of Sciences* 98(5):2170-2175.
5. Sharkey TD (1988) Estimating the rate of photorespiration in leaves. *Physiologia plantarum* 73(1):147-152.
6. Bauwe H, Hagemann M, & Fernie AR (2010) Photorespiration: players, partners and origin. *Trends in plant science* 15(6):330-336.
7. Sage RF, Sage TL, & Kocacinar F (2012) Photorespiration and the evolution of C4 photosynthesis. *Annual review of plant biology* 63:19-47.
8. Giordano M, Beardall J, & Raven JA (2005) CO_2 concentrating mechanisms in algae: mechanisms, environmental modulation, and evolution. *Annu. Rev. Plant Biol.* 56:99-131.
9. Behrenfeld MJ*, et al.* (2001) Biospheric primary production during an ENSO transition. *Science* 291(5513):2594-2597.
10. Rousseaux CS & Gregg WW (2013) Interannual variation in phytoplankton primary production at a global scale. *Remote Sensing* 6(1):1-19.
11. Mann GD (1996) Chloroplast morphology, movements and inheritance in diatoms. *Cytology, genetics and molecular biology of algae*, eds Chaudhary BR & Agrawal SB (SPB Academic Publishing, Amsterdam), pp 249-274.
12. Thierstein HR & Young JR (2004) *Coccolithophores: from molecular processes to global impact* (Springer Science & Business Media).
13. Meyer M & Griffiths H (2013) Origins and diversity of eukaryotic CO_2-concentrating mechanisms: lessons for the future. *Journal of experimental botany* 64(3):769-786.
14. McKay RML & Gibbs SP (1991) Composition and function of pyrenoids: cytochemical and immunocytochemical approaches. *Canadian Journal of Botany* 69(5):1040-1052.
15. Schmitz F (1882) *Die Chromatophoren der Algen: Vergleichende Untersuchungen über Bau und Entwicklung der Chlorophyllkörper und der analogen Farbstoffkörper der Algen* (M. Cohen & Sohn (F. Cohen)).
16. Vaucher J-P (1803) *Histoire des conferves d'eau douce: contenant leurs différens modes de reproduction, et la description de leurs principales espèces, suivie de l'histoire des trémelles et des ulves d'eau douce* (JJ Paschoud).
17. Brown R (1967) Pyrenoid: its structure distribution and function. *Journal of Phycology* 3(Suppl. 1):5-7.
18. Berry J, Boynton J, Kaplan A, & Badger M (1976) Growth and photosynthesis of *Chlamydomonas reinhardtii* as a function of CO_2 concentration. *Carnegie Institution of Washington Year Book* 75:423-432
19. Borkhsenious ON, Mason CB, & Moroney JV (1998) The intracellular localization of ribulose-1, 5-bisphosphate carboxylase/oxygenase in *Chlamydomonas reinhardtii*. *Plant physiology* 116(4):1585-1591.
20. Miura K*, et al.* (2004) Expression profiling-based identification of CO_2-responsive genes regulated by CCM1 controlling a carbon-concentrating mechanism in *Chlamydomonas reinhardtii*. *Plant physiology* 135(3):1595-1607.
21. Schwanhäusser B*, et al.* (2011) Global quantification of mammalian gene expression control. *Nature* 473(7347):337-342.
22. Turkina MV, Blanco-Rivero A, Vainonen JP, Vener AV, & Villarejo A (2006) CO_2 limitation induces specific redox-dependent protein phosphorylation in *Chlamydomonas reinhardtii*. *Proteomics* 6(9):2693-2704.
23. Dunker AK, Silman I, Uversky VN, & Sussman JL (2008) Function and structure of inherently disordered proteins. *Current opinion in structural biology* 18(6):756-764.
24. Engel BD*, et al.* (2015) Native architecture of the *Chlamydomonas* chloroplast revealed by in situ cryo-electron tomography. *eLife* 4:DOI: 10.7554/eLife.04889.
25. Meyer MT*, et al.* (2012) Rubisco small-subunit α-helices control pyrenoid formation in *Chlamydomonas*. *Proceedings of the National Academy of Sciences* 109(47):19474-19479.
26. Long BM, Badger MR, Whitney SM, & Price GD (2007) Analysis of carboxysomes from *Synechococcus* PCC7942 reveals multiple Rubisco complexes with carboxysomal proteins CcmM and CcaA. *Journal of Biological Chemistry* 282(40):29323-29335.
27. Villarreal JC & Renner SS (2012) Hornwort pyrenoids, carbon-concentrating structures, evolved and were lost at least five times during the last 100 million years. *Proceedings of the National Academy of Sciences* 109(46):18873-18878.
28. Mitchell MC, Meyer MT, & Griffiths H (2014) Dynamics of carbon-concentrating mechanism induction and protein relocalization during the dark-to-light transition in synchronized *Chlamydomonas reinhardtii*. *Plant physiology* 166(2):1073-1082.
29. Wang H*, et al.* (2014) The global phosphoproteome of *Chlamydomonas reinhardtii* reveals complex organellar phosphorylation in the flagella and thylakoid membrane. *Molecular & Cellular Proteomics* 13(9):2337-2353.
30. Kay BK, Williamson MP, & Sudol M (2000) The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *The FASEB journal* 14(2):231-241.

31. Ma Y, Pollock SV, Xiao Y, Cunnusamy K, & Moroney JV (2011) Identification of a novel gene, CIA6, required for normal pyrenoid formation in *Chlamydomonas reinhardtii*. *Plant physiology* 156(2):884-896.
32. Long SP, Marshall-Colon A, & Zhu XG (2015) Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell* 161(1):56-66.
33. Sager R (1955) Inheritance in the Green Alga *Chlamydomonas Reinhardi*. *Genetics* 40(4):476-489.
34. Gorman DS & Levine R (1965) Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardi*. *Proceedings of the National Academy of Sciences of the United States of America* 54(6):1665-1669.
35. Khrebtukova I & Spreitzer RJ (1996) Elimination of the *Chlamydomonas* gene family that encodes the small subunit of ribulose-1, 5-bisphosphate carboxylase/oxygenase. *Proceedings of the National Academy of Sciences* 93(24):13689-13693.
36. Zhang R*, et al.* (2014) High-Throughput Genotyping of Green Algal Mutants Reveals Random Distribution of Mutagenic Insertion Sites and Endonucleolytic Cleavage of Transforming DNA. *The Plant cell* 26(4):1398-1409.
37. Kropat J*, et al.* (2011) A revised mineral nutrient supplement increases biomass and growth rate in *Chlamydomonas reinhardtii*. *The Plant journal* 66(5):770-780.
38. Spreitzer RJ & Mets L (1981) Photosynthesis-deficient Mutants of *Chlamydomonas reinhardii* with Associated Light-sensitive Phenotypes. *Plant physiology* 67(3):565-569.
39. Mettler T*, et al.* (2014) Systems Analysis of the Response of Photosynthesis, Metabolism, and Growth to an Increase in Irradiance in the Photosynthetic Model Organism *Chlamydomonas reinhardtii*. *The Plant cell* 26(6):2310-2350.
40. Brueggeman AJ*, et al.* (2012) Activation of the carbon concentrating mechanism by CO_2 deprivation coincides with massive transcriptional restructuring in *Chlamydomonas reinhardtii*. *The Plant cell* 24(5):1860-1875.
41. Fang W*, et al.* (2012) Transcriptome-wide changes in *Chlamydomonas reinhardtii* gene expression regulated by carbon dioxide and the CO_2-concentrating mechanism regulator CIA5/CCM1. *The Plant cell* 24(5):1876-1893.
42. Nordhues A*, et al.* (2012) Evidence for a role of VIPP1 in the structural organization of the photosynthetic apparatus in *Chlamydomonas*. *The Plant cell* 24(2):637-659.
43. Kuchitsu K, Tsuzuki M, & Miyachi S (1988) Changes of Starch Localization within the Chloroplast Induced by Changes in CO_2 Concentration during Growth of *Chlamydomonas reinhardtii*: Independent Regulation of Pyrenoid Starch and Stroma Starch. *Plant and Cell Physiology* 29(8):1269-1278.
44. Kuchitsu K, Tsuzuki M, & Miyachi S (1991) Polypeptide composition and enzyme activities of the pyrenoid and its regulation by CO_2 concentration in unicellular green algae. *Canadian Journal of Botany* 69(5):1062-1069.
45. Lowry OH, Rosebrough NJ, Farr AL, & Randall RJ (Protein measurement with the Folin phenol reagent.
46. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227(5259):680-685.
47. Mühlhaus T, Weiss J, Hemme D, Sommer F, & Schroda M (2011) Quantitative shotgun proteomics using a uniform ^{15}N-labeled standard to monitor proteome dynamics in time course experiments reveals new insights into the heat stress response of *Chlamydomonas reinhardtii*. *Molecular & Cellular Proteomics* 10(9):M110. 004739.
48. Cox J & Mann M (2008) MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* 26(12):1367-1372.
49. Gibson DG*, et al.* (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* 6(5):343-345.
50. Yang W*, et al.* (2014) Alternative acetate production pathways in *Chlamydomonas reinhardtii* during dark anoxia and the dominant role of chloroplasts in fermentative acetate production. *The Plant cell* 26(11):4499-4518.
51. Livak KJ & Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25(4):402-408.
52. Schloss JA (1990) A *Chlamydomonas* gene encodes a G protein β subunit-like polypeptide. *Molecular and General Genetics MGG* 221(3):443-452.
53. Heinnickel ML*, et al.* (2013) Novel thylakoid membrane GreenCut protein CPLD38 impacts accumulation of the cytochrome b6f complex and associated regulatory processes. *The Journal of biological chemistry* 288(10):7024-7036.
54. Wellburn AR (1994) The Spectral Determination of Chlorophylls a and b, as well as Total Carotenoids, Using Various Solvents with Spectrophotometers of Different Resolution. *Journal of plant physiology* 144(3):307-313.
55. Badger MR, Kaplan A, & Berry JA (1980) internal inorganic carbon pool of *Chlamydomonas reinhardtii* Evidence for a carbon dioxide-concentrating mechanism. *Plant physiology* 66(3):407-413.
56. Heuser JE (2011) The origins and evolution of freeze-etch electron microscopy. *Journal of Electron Microscopy* 60(suppl 1):S3-S29.
57. Skepper J (2000) Immunocytochemical strategies for electron microscopy: choice or compromise. *Journal of microscopy* 199(1):1-36.
58. Obradovic Z*, et al.* (2003) Predicting intrinsic disorder from amino acid sequence. *Proteins: Structure, Function, and Bioinformatics* 53(S6):566-572.
59. Linding R, Russell RB, Neduva V, & Gibson TJ (2003) GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic acids research* 31(13):3701-3708.

60. Buchan DW, Minneci F, Nugent TC, Bryson K, & Jones DT (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic acids research* 41(W1):W349-W357.
61. Kelley LA, Mezulis S, Yates CM, Wass MN, & Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* 10(6):845-858.
62. Newman AM & Cooper JB (2007) XSTREAM: a practical algorithm for identification and architecture modeling of tandem repeats in protein sequences. *BMC bioinformatics* 8(1):382.
63. Romero P*, et al.* (2001) Sequence complexity of disordered protein. *Proteins: Structure, Function, and Bioinformatics* 42(1):38-48.
64. McDonald JH (2009) *Handbook of biological statistics* (Sparky House Publishing Baltimore, MD).
65. Flombaum P*, et al.* (2013) Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proceedings of the National Academy of Sciences* 110(24):9824-9829.
66. Not F*, et al.* (2004) A single species, *Micromonas pusilla* (Prasinophyceae), dominates the eukaryotic picoplankton in the Western English Channel. *Applied and Environmental Microbiology* 70(7):4064-4072.
67. Chrétiennot-Dinet M*, et al.* (1995) A new marine picoeucaryote: *Ostreococcus tauri* gen. et sp. nov.(Chlorophyta, Prasinophyceae). *Phycologia* 34(4):285-292.

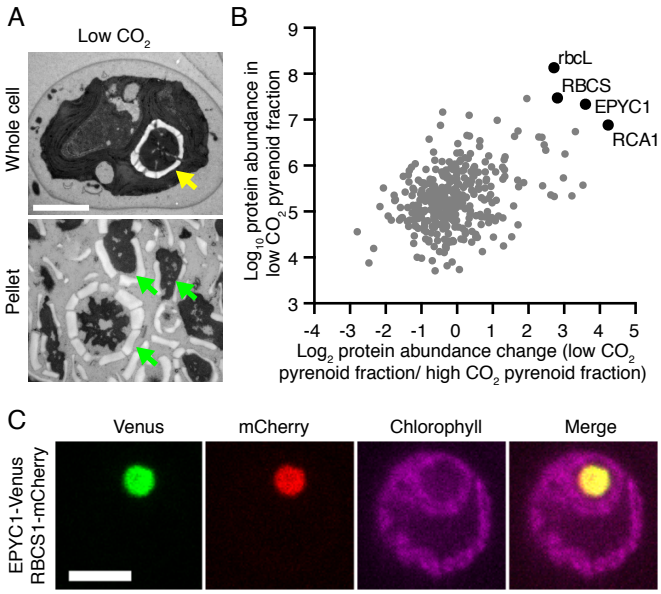


Fig. 1. EPYC1 is an abundant pyrenoid protein. (*A*), TEM images of *Chlamydomonas reinhardtii* whole cells and pyrenoid-enriched pellet fraction from cells grown at low CO_2. Yellow arrow indicates the pyrenoid; green arrows indicate pyrenoid-like structures. Scale bar: 2 µm. (*B*) Mass-spectrometric analysis of 366 proteins in pyrenoid-enriched pellet fractions from low- and high-CO_2-grown cells (mean of 4 biological replicates; for raw data, see Dataset 1). RbcL, RBCS, EPYC1 and RCA1 (black) are abundant in low CO_2 pellets (determined by intensity-based absolute quantification (iBAQ); y-axis). Additionally, these proteins showed increased abundance in low CO_2 compared to high CO_2 pellets (determined by label-free quantification (LFQ); x-axis). (*C*) Confocal microscopy of EPYC1-Venus and RBCS1-mCherry co-expressed in wild-type cells. Scale bar: 5µm.

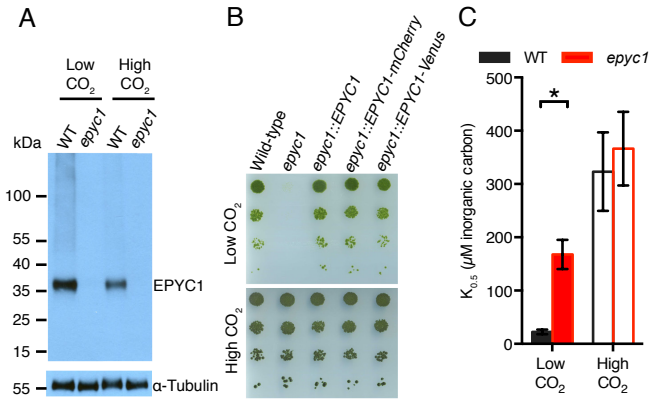


Fig. 2. EPYC1 is an essential component of the carbon concentrating mechanism. (*A*) EPYC1 protein levels in WT and *epyc1* mutant cells grown at low and high CO_2 were probed by western blotting with anti-EPYC1 antibodies. Anti-tubulin is shown as a loading control. (*B*) Growth phenotypes of WT, *epyc1* and three lines complemented with *EPYC1*. Serial 1:10 dilutions of WT, *epyc1*, *eypc*::*EPYC1*, *epyc1*::*EPYC1-mCherry* and *epyc1*::*EPYC1-Venus* lines were spotted on TP minimal medium and grown at low and high CO_2 under 500 µmol photons m^{-2} s^{-1} illumination. (*C*) Inorganic carbon affinity of wild-type (WT) and *epyc1* cells. Cells were pre-grown at low or high CO_2, and whole-cell inorganic carbon affinity was measured as the concentration of inorganic carbon at half maximal O_2 evolution (data is a mean of 5 low CO_2 or 3 high CO_2 biological replicates; error bars: SEM; asterisk: P=0.0055, Student's t-test).

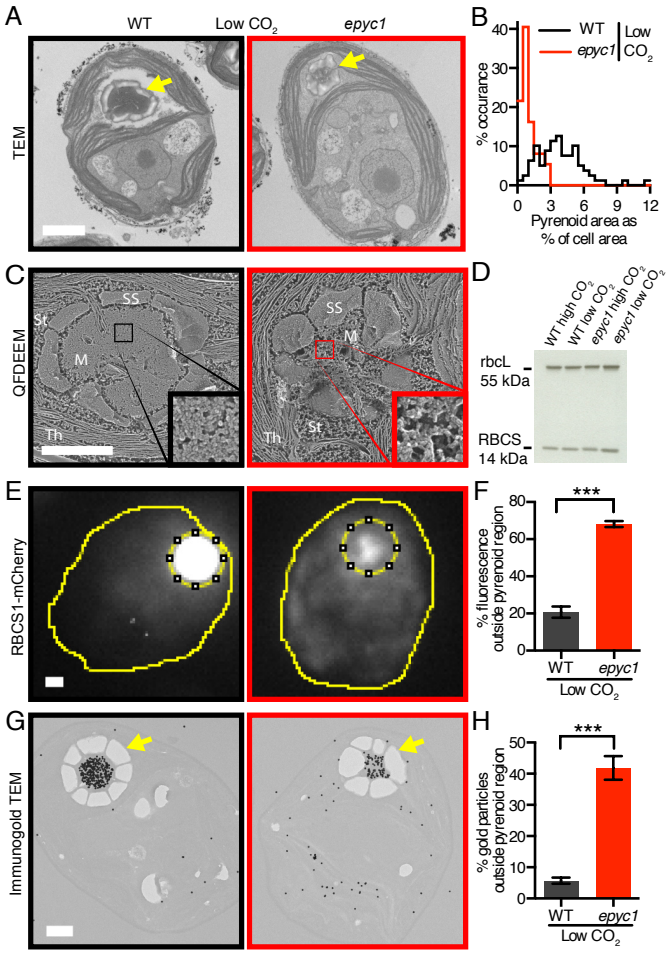


Fig. 3. EPYC1 is essential for Rubisco aggregation in the pyrenoid. (*A*) Representative TEMs of WT and *epyc1* cells grown at low CO_2. (*B*) Quantification of pyrenoid area as percentage of cell area of WT and *epyc1* cells grown at low CO_2 (data is from TEM images as represented in (A), *epyc1*: n=37, WT: n=79, $P<10^{-19}$, Welch's *t*-test). (*C*) Quick-freeze deep-etch electron microscopy (QFDEEM) of the pyrenoid of WT and *epyc1* cells grown at low CO_2. M, pyrenoid matrix; St, stroma; Th, thylakoids; SS, starch sheath. Insets are a 400% zoom of the pyrenoid matrix. (*D*) Rubisco protein levels in WT and *epyc1* cells grown at low and high CO_2 were probed by western blotting. (*E*) The localization of Rubisco was determined by microscopy of WT and *epyc1* mutants containing RBCS1-mCherry. The sum of fluorescence signal from Z stacks is shown and was used for quantitation. (*F*) The fraction of RBCS1-mCherry signal from outside the pyrenoid region (inner dotted line, *E)* was quantified in WT and *epyc1* cells at low CO_2 (*epyc1*: n=27, WT: n=27, *** represents $P<10^{-15}$, Student's *t*-test). (*G*) Representative images of anti-Rubisco immunogold labeling of WT and *epyc1* cells grown at low CO_2. Gold particles were enlarged 10x for visibility. (*H*) The fraction of immunogold particles outside the pyrenoid was quantified (WT: n= 26 cells, 8123 gold particles; *epyc1*: n=27 cells, 2708 gold particles; *** represents $P<10^{-15}$, Student's *t*-test). (*F*) and (*H*): mean values with error bars indicating SEM Yellow arrows indicate pyrenoids. Scale bars: 1 µm.

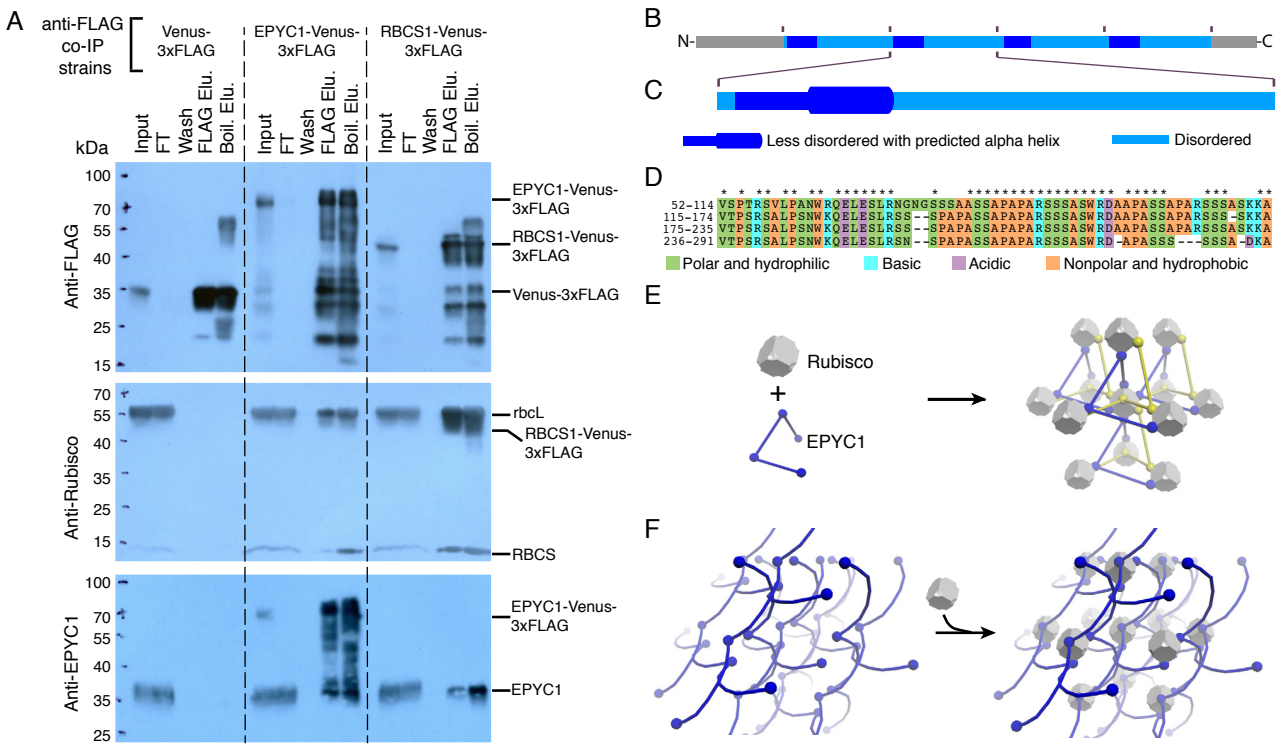


Fig. 4. EPYC1 forms a complex with Rubisco. (*A*) Anti-FLAG co-immunoprecipitations (co-IPs) of WT cells expressing Venus-3xFLAG, EPYC1-Venus-3xFLAG and RBCS1-Venus-3xFLAG are shown. For each co-IP, the input, flow-through (FT), 4th wash (wash), 3xFLAG elution (FLAG Elu.) and boiling elution (Boil. Elu.) were probed with anti-FLAG, anti-Rubisco or anti-EPYC1. Right hand side labels show the expected sizes of proteins. **(***B*) Analysis of the EPYC1 protein sequence shows that EPYC1 consists of four nearly identical repeats. (*C*) Each repeat has a highly disordered domain (light blue) and a less disordered domain (dark blue) containing a predicted alpha-helix (thicker line) rich in charged residues. (*D*) Amino acid alignments of the four repeats are shown. Asterisks indicate residues that are identical in all four repeats. (*E* and *F*) Two models illustrate how EPYC1 could bind the Rubisco holoenzyme in a manner that is compatible with the observed packing of Rubisco in the pyrenoid. (*E*) EPYC1 and Rubisco could form a co-dependent network. If each EPYC1 can bind four Rubisco holoenzymes, and each Rubisco holoenzyme can bind eight EPYC1s, eight EPYC1 proteins could connect each Rubisco to twelve neighboring Rubiscos. (*F*) EPYC1 could form a scaffold onto which Rubisco binds. Both arrangements could expand indefinitely in every direction. For clarity, the spacing between Rubisco holoenzymes was increased and EPYC1 is depicted in both yellow and blue.

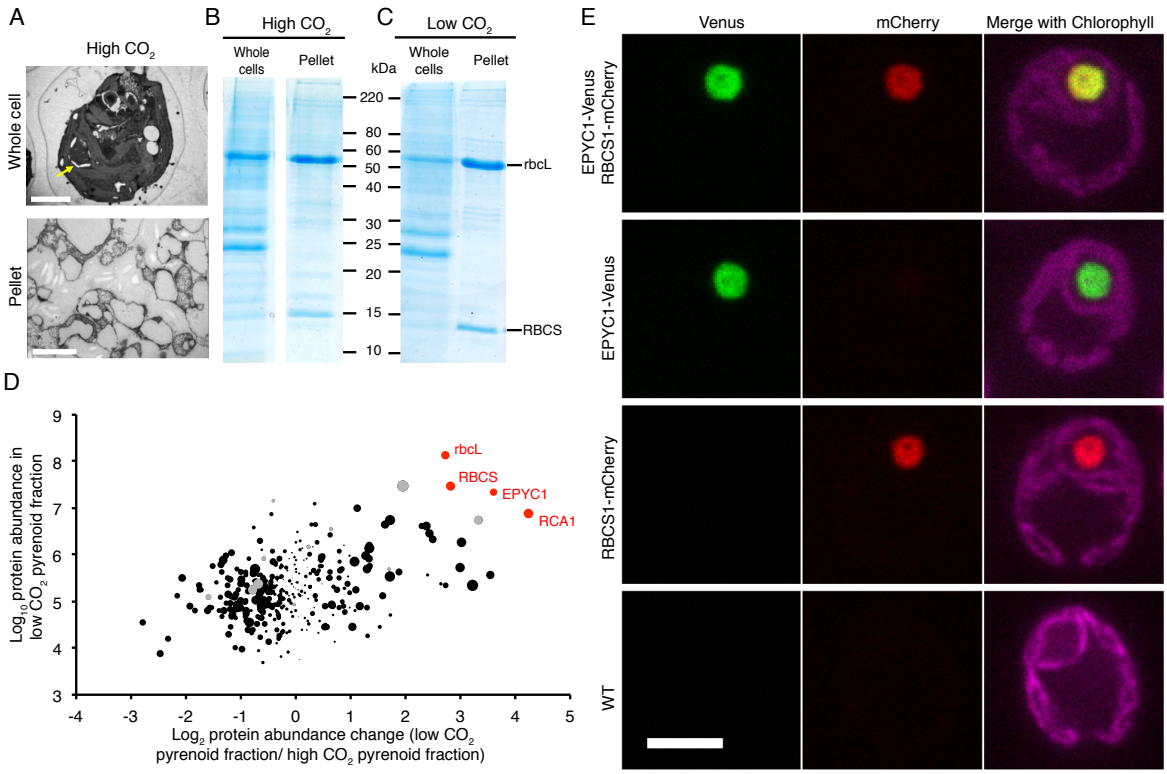


Fig. S1. EPYC1 is an abundant pyrenoid component. (*A*) TEM images of *Chlamydomonas reinhardtii* whole cells and pyrenoid-enriched pellet from cells grown at high CO_2. Yellow arrow indicates pyrenoid. Scale bars: 2 µm. *(B)* Coomassie-stained SDS-PAGE of whole cell and pyrenoid-enriched pellet at high CO_2. (*C*) Coomassie-stained SDS-PAGE of whole cell and pyrenoid-enriched pellet at low CO_2. (*D*) Additional analysis of the mass spectrometry data shown in Fig. 1*B* and Dataset S1. The x-axis is the label-free quantification (LFQ) enrichment in the low-CO_2 pellet fraction relative to the high CO_2 pellet fraction. The iBAQ given on the y-axis, represents the absolute protein abundance in the low-CO_2 pellet. Red data points highlight RbcL, RBCS, EPYC1 and RCA1. Grey circles depict sets of peptides represented by more than one protein due to high sequence similarities, whereas black circles are peptides representing a single protein. Dot sizes indicate the $\log_{10}$ *P*-value between low CO_2 and high CO_2 pellet fractions (Student´s *t*-test). In total 366 proteins were identified in all four replicates of both the low- and high-CO_2-grown pellets. (*E*) Confocal microscopy of EPYC1-Venus and RBCS1-mCherry co-expressed and individually expressed in WT cells. All images were taken and processed with the same settings. Top panel is the same cell as shown in Fig. 1*C*. Scale bar: 5µm.

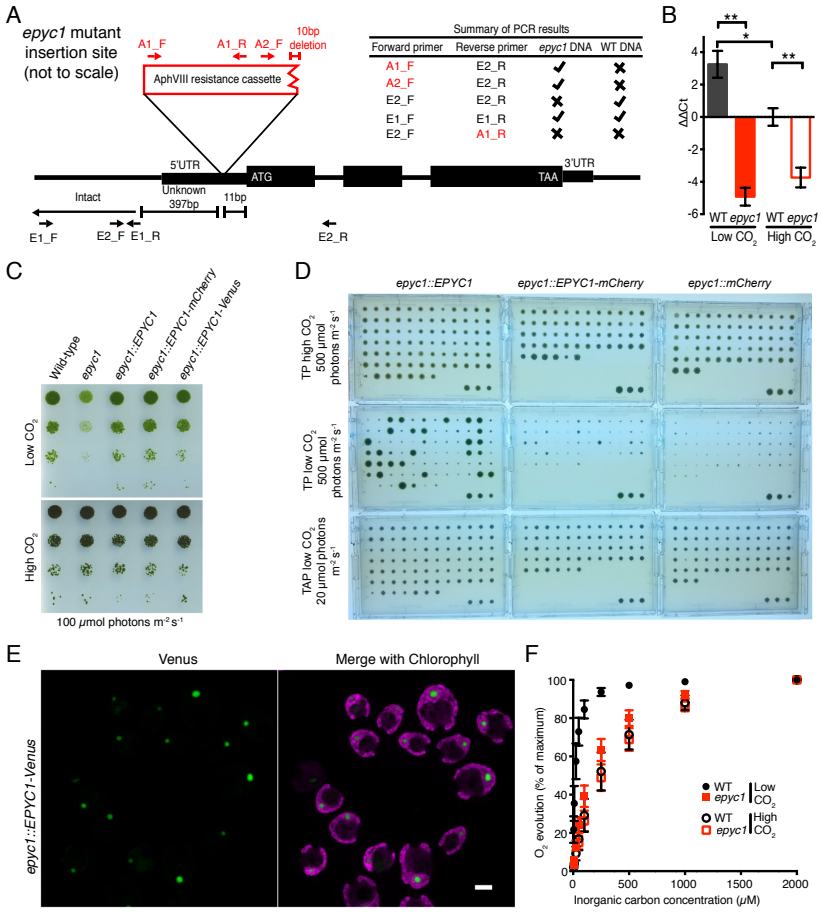


Fig. S2. Characterization and complementation of the *epyc1* mutant. (*A*) Cartoon of the *epyc1* mutant insertion site. The pMJ016c resistance cassette conferring paromomycin through the AphVIII gene is 11bp upstream of the *EPYC1* ATG start codon. The resistance cassette has a known 10bp deletion at the 3' end and is fully intact. The junction of the 3' cassette end and the *EPYC1* gene can be amplified with a forward primer annealing at the 5' end of the cassette (A1_F) and a reverse primer in the *EPYC1* gene (E2_R). The 5' end of the insert is still uncharacterized. Primers upstream of the insertion site fail to give PCR products when paired with reverse primers in the resistance cassette (e.g. E2_F and A1_R). It is known that insertion sites can undergo large insertions and/or deletions in *Chlamydomonas* (36). However, a large deletion upstream of the insertion site is ruled out due to the amplification of a region upstream of the insertion site (E1_F and E1_R) in the *epyc1* mutant. Note the cartoon is not to scale for clarity. (*B*) Quantification of *EPYC1* transcript levels in WT and the *epyc1* mutant at low and high CO_2 by qRT-PCR. Transcript levels are normalized to the reference gene *RCK1*, and plotted relative to WT at high CO_2. In the *epyc1* mutant, transcript levels were ~250-fold lower than in WT at low CO_2; and transcript levels were not significantly upregulated between low and high CO_2 (P=0.129, Student's t-test). Data is the mean of 3 biological replicates each measured in triplicate. Error bars: SEM * indicates $P< 0.05$, ** indicates $P< 0.005$, Student's t-test. (*C*) Growth phenotypes of WT, *epyc1* and 3 *epyc1* complemented lines at 100 µmol photons m^{-2} s^{-1} light intensity. Serial 1:10 dilutions of WT, *epyc1*, *eypc*::*EPYC1*, *epyc1*::*EPYC1-mCherry* and *epyc1*::*EPYC1-Venus* lines were spotted on TP minimal plates and grown at low and high CO_2 under 100 µmol photons m^{-2} s^{-1}. (*D*) Complementation screening of the *epyc1* mutant. The *epyc1* mutant was transformed with pLM006_EPYC1, pLM006_EPYC1-mCherry or pLM006_mCherry and selected on TAP plates with hygromycin. Hygromycin resistant colonies were picked into a 96 format and propagated twice on TAP with hygromycin plates to allow even growth of all colonies. Colonies were then replicated onto TP or TAP plates and incubated as shown for 12 days before imaging. pLM006_EPYC1 fully rescued the *epyc1* mutant in 28% (22/79) of cases and partially rescued the mutant in 10% (8/79) of cases. pLM006_EPYC1-mCherry partially rescued the mutant in 20% (11/54) of cases. The negative control, pLM006_mCherry, failed to rescue the mutant (0/62). Partially rescued colonies were colonies that were visibly smaller in size than WT but larger in size than negative control colonies. The residual growth seen in all colonies at low CO_2 is carryover from the initial pinning from TAP plates. The 3 bottom right colonies are WT controls. (*E*) Confocal microscopy of the *epyc1::EPYC1-Venus* complemented line used for spot tests in Fig. 2*B*. Cells were grown mixotrophically in TAP media and imaged by confocal microscopy as in the materials and methods section. Scale bar: 5 µm. (*F*) The *epyc1* mutant has reduced inorganic carbon affinity. WT and *epyc1* cells were grown at low and high CO_2, and whole cell inorganic carbon affinity was measured by O_2 evolution with step-wise increases in inorganic carbon. The $K_{0.5}$ values shown in Fig. 2*C* are derived from these curves. Data is a mean of 5 low CO_2 or 3 high CO_2 biological replicates. Error bars: SD.

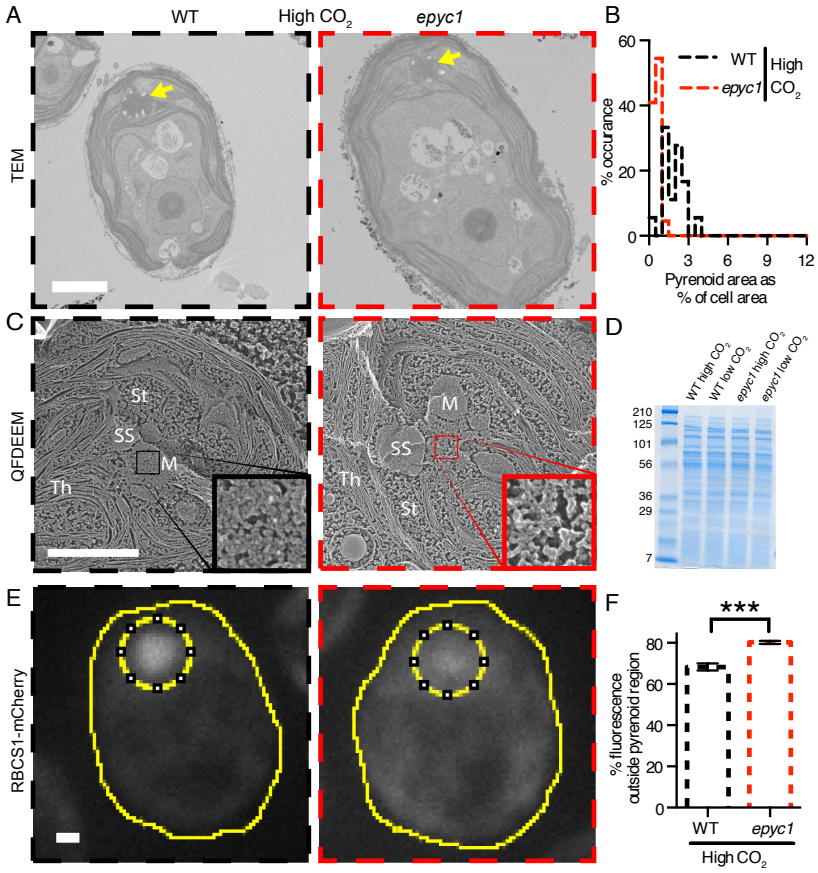


Fig. S3. Rubisco is mislocalized in the *epyc1* mutant at high CO_2. (*A*) Representative TEMs of WT and *epyc1* cells grown at high CO_2. Yellow arrows indicate pyrenoids. (*B*) Quantification of pyrenoid area as percentage of cell area of WT and *epyc1* cells grown at high CO_2 (WT: n=18, *epyc1*: n=22, $P<10^{-5}$, Welch's *t*-test). (*C*) Quick-Freeze Deep-Etch EM (QFDEEM) of the pyrenoid of WT and *epyc1* cells grown at high CO_2. M, pyrenoid matrix; St, stroma; Th, thylakoids; SS, starch sheath. Insets are a 400% zoom of the pyrenoid matrix. (*D*) Coomassie stained SDS-PAGE loading control gel for samples used in Fig. 3*D*. (*E*) The localization of Rubisco was determined by microscopy of WT and *epyc1* mutants containing RBCS1-mCherry at high CO_2. (*F*) The fraction of RBCS1-mCherry signal from outside the pyrenoid region (inner dotted line, *E*) was quantified in WT and *epyc1*. The sum of fluorescence signal from Z stacks is shown and was used for quantitation. WT: n=20, *epyc1*: n=20, *** represents $P=10^{-6}$, Student's *t*-test. Scale bars: 1 µm.

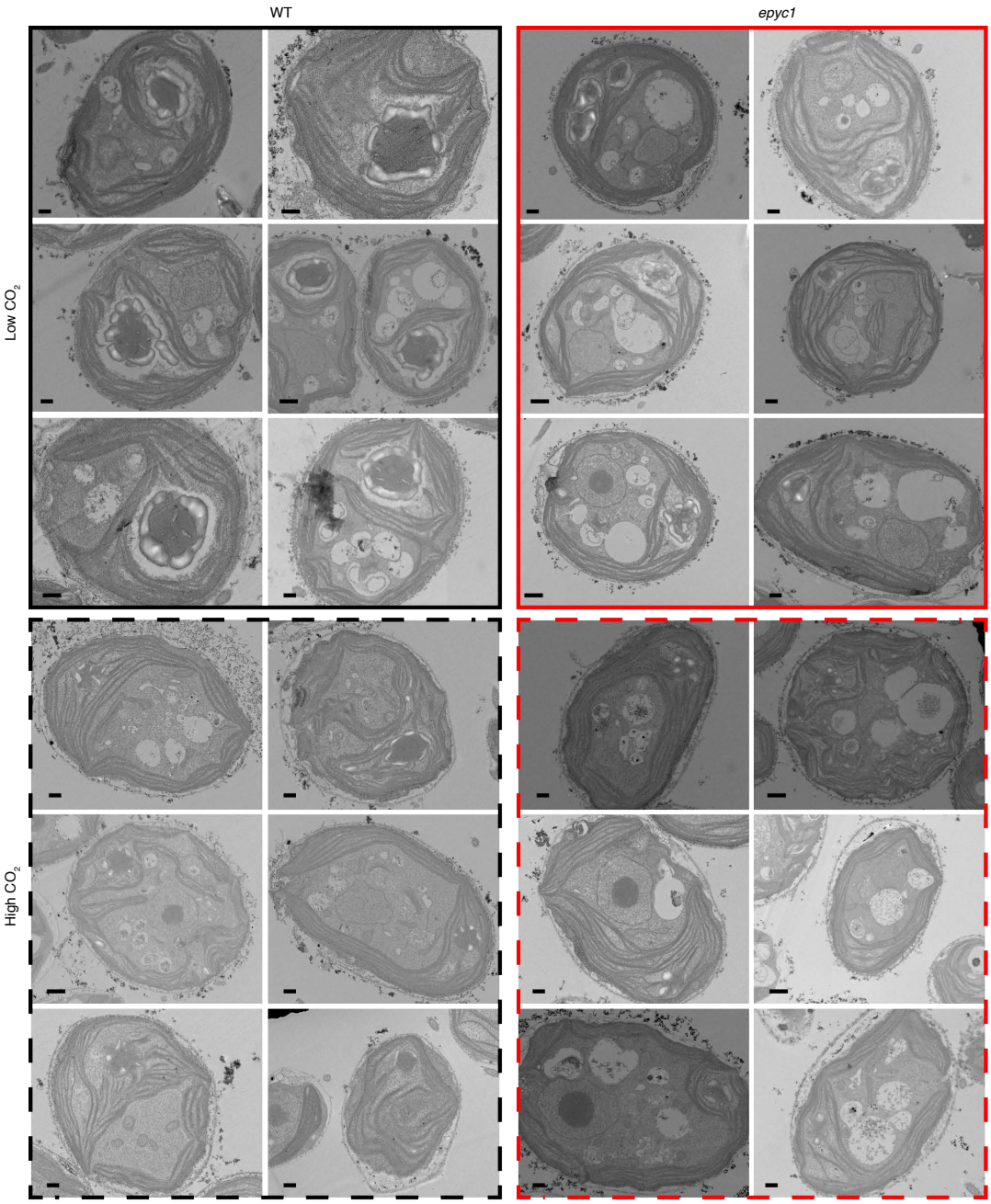


Fig. S4. Representative TEM images of WT and *epyc1* cells at low and high CO_2. A representative selection of TEM images used for pyrenoid area analysis. Cells were prepared and imaged for TEM as in the materials and methods. Scale bars, 500 nm.

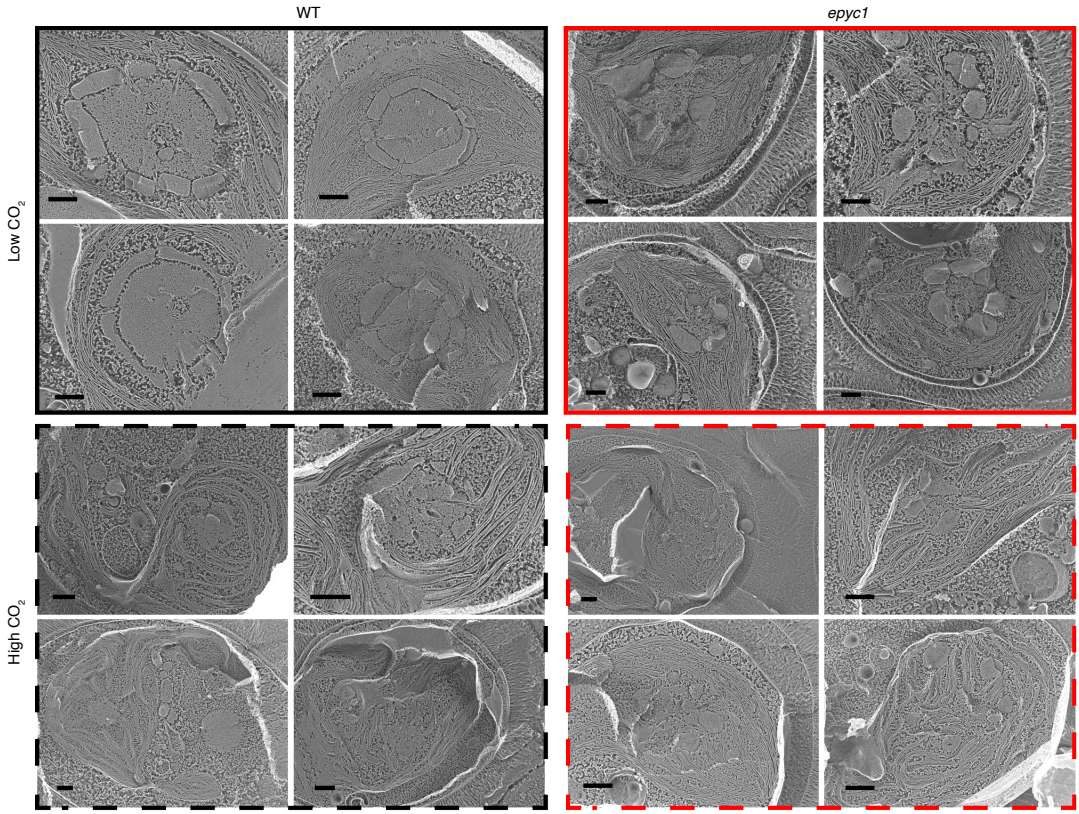


Fig. S5. Representative Quick-Freeze Deep-Etch EM (QFDEEM) images of WT and *epyc1* cells at low and high CO_2. Cells were prepared and imaged for QFDEEM as in the materials and methods. Scale bars: 500 nm.

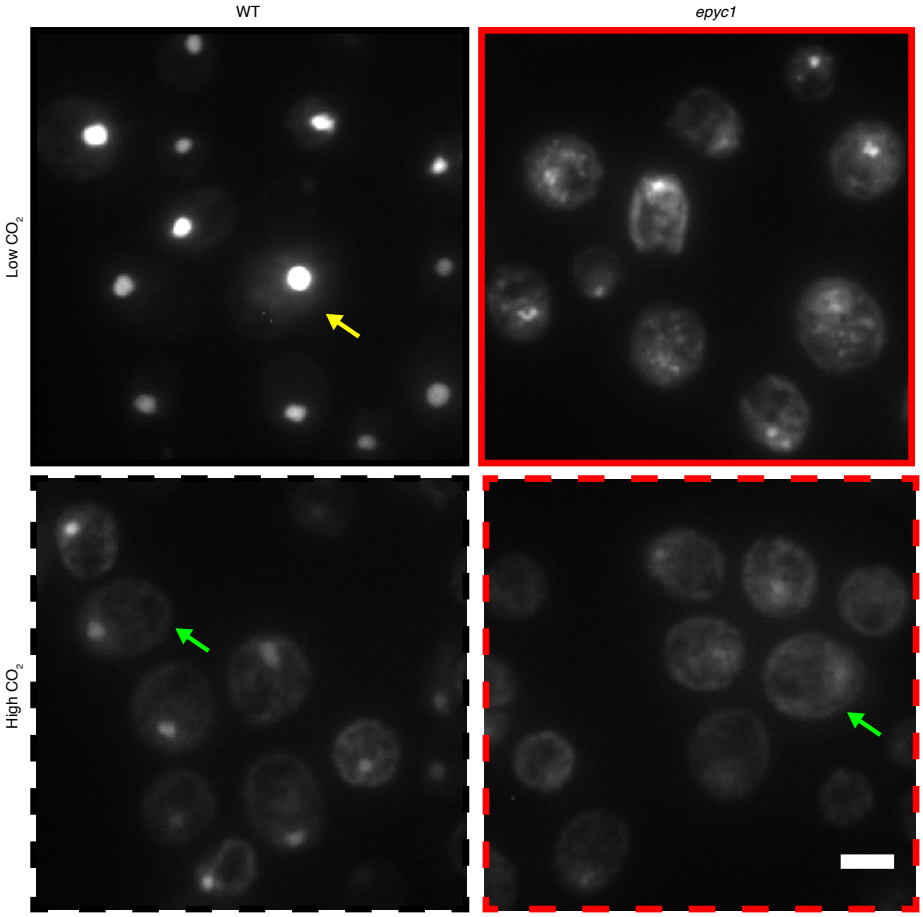


Fig. S6. Representative cells used for Rubisco-mCherry localization data. A representative field of view used for quantifying the mislocalization of Rubisco in the *epyc1* mutant. Images are summed z-stacks of 40 confocal sections 0.3 µm apart. Yellow arrow indicates the cell used for Fig. 3*E*. Green arrows indicate cells used for Fig. S3*E*. Scale bar: 5 µm.

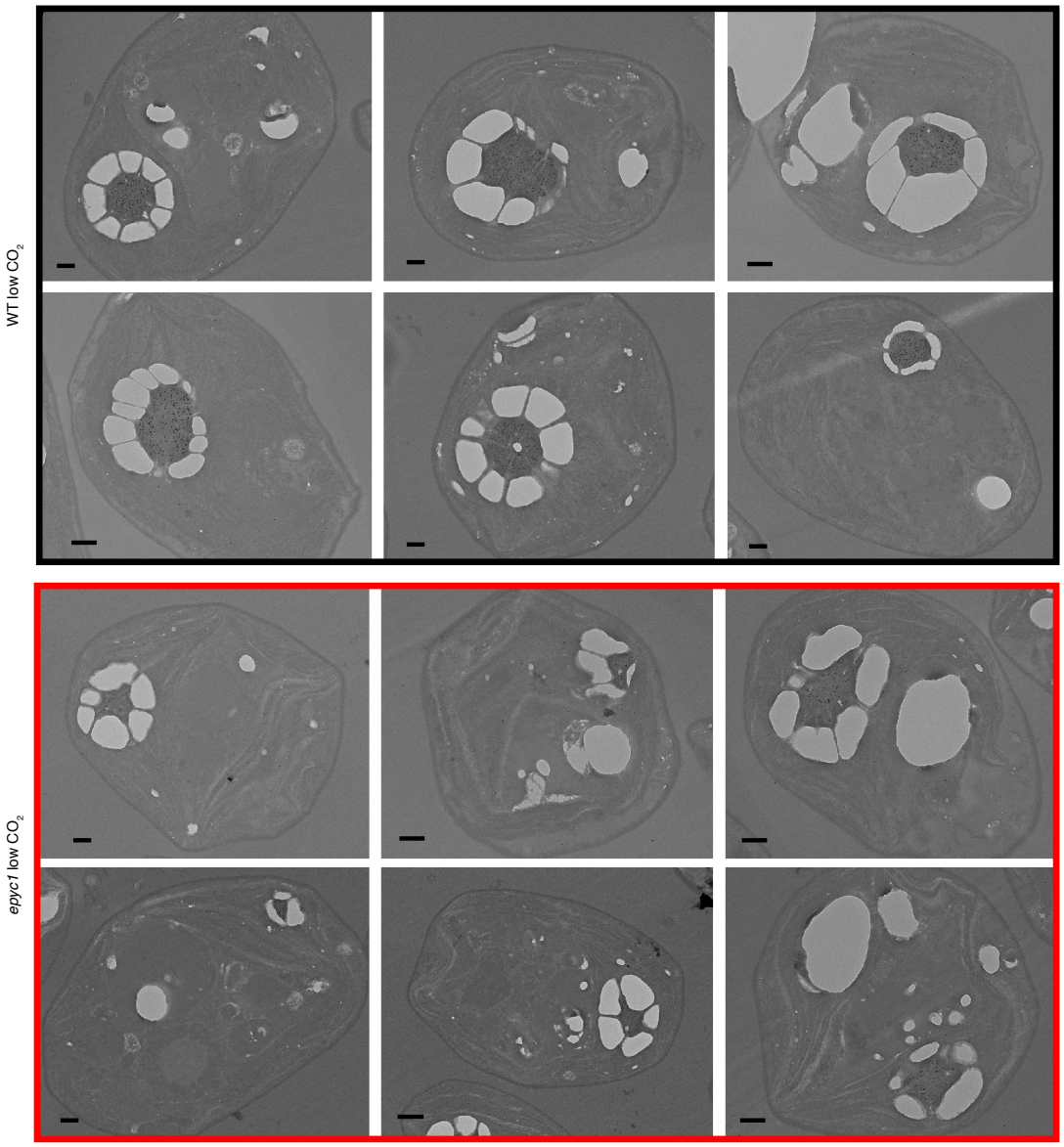


Fig. S7. Representative cells used for Rubisco immunogold labeling. A representative selection of immunogold-TEM images used for quantification of Rubisco levels outside the pyrenoid. The top left cell for each condition is the cell used in Fig. 3*G* before gold particle enlargement. Cells were prepared and imaged for immunogold-TEM as in the materials and methods. Scale bars: 500 nm.

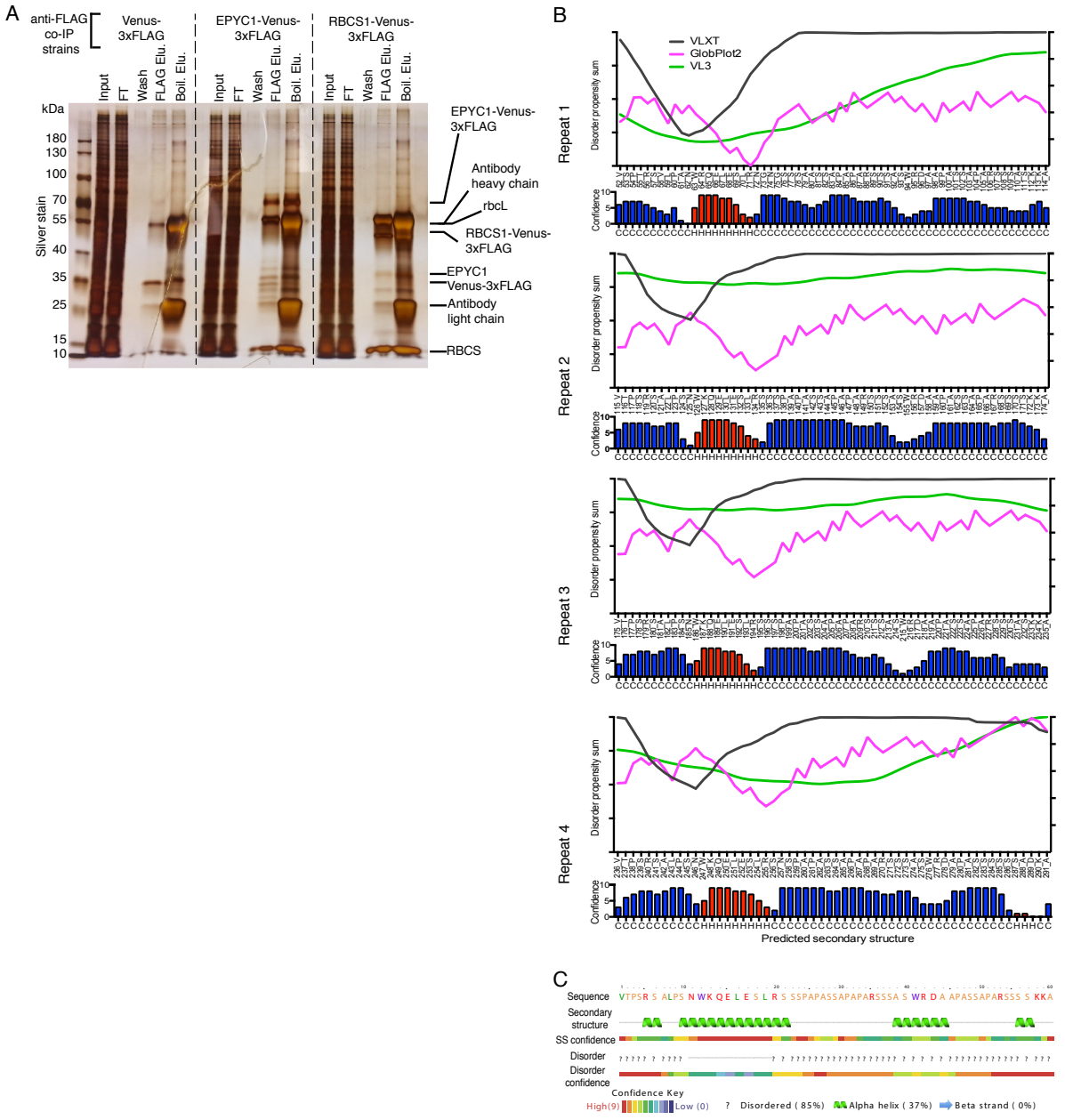


Fig. S8. EPYC1 interactions and sequence analysis. (*A*) Anti-FLAG co-immunoprecipitations (co-IPs) of WT cells expressing Venus-3xFLAG, EPYC1-Venus-3xFLAG and RBCS1-Venus-3xFLAG are shown. For each co-IP, the input, flow-through (FT), 4th wash (wash), 3xFLAG elution (FLAG Elu.) and boiling elution (Boil. Elu.) were run on an SDS-PAGE gel and silver stained. Right hand side labels show the expected sizes of proteins. (*B*) and (*C*) Analysis of the EPYC1 protein sequence. (*B*) To investigate the disorder of EPYC1, the full-length amino acid sequence was analyzed by VL3, VLTX and GlobPlot2 disorder prediction algorithms. The lower bar chart shows the PSIPRED v3.3 predicted secondary structure of full-length EPYC1, H = helix (red), C = coil (blue). Bar height indicates confidence value. (*C*) Analysis of the repeat region from 115-174 by Phyre2.

Table S1. The contribution of the pyrenoid to global net primary production

	Percentage of ocean NPP	Percentage of algal group with a pyrenoid	Percentage of ocean NPP mediated by a pyrenoid	Percentage of total NPP
Global primary production				100%
Terrestrial				50 (9) - 54% (3)
Ocean				46 (3) - 50% (9)
Cyanobacteria (Prokaryotic)	10 (10) - 25% (65)			
Eukaryotic algae	75 (65) - 90% (10)			
Diatoms	42 (10, 65) - 50% (10)	100% (11)	42-50%	
Coccolithophores	17 (10, 65) - 20% (10)	100% (12)	17-20%	
Chlorophytes	17 (10, 65) - 20% (10)	10-90%*	2-18%	
Pyrenoid containing algae			61-88%	28 – 44%

*The majority of chlorophytes are known to have pyrenoids (13), with the pyrenoid containing *Micromonas pusilla* shown to be the dominant species in several oceanic and coastal regions (66). However, some ocean chlorophytes, including the abundant species *Bathycoccus prasinos,* appear to lack pyrenoids (13).

Table S2. Oligonucleotides used in this work

Primer name	Sequence
EPYC1_ORF_F	GCTACTCACAACAAGCCCAGTTATGGCCACTATCTCGTCGATGCGC
EPYC1_ORF_R	GAGCCACCCAGATCTCCGTTCAGGCCCTTGCGCCAGTCAGC
RBCS1_ORF_F	GCTACTCACAACAAGCCCAGTTATGGCCGCCGTCATTGCCAAGTC
RBCS1_ORF_R	GAGCCACCCAGATCTCCGTTCACGGAGCGCTTGTTGGCGGG
GBLP_F:	AACACCGTGACCGTCTCC
GBLP_R:	TGCTGGTGATGTTGAACTCG
EPYC1_F:	AAGCAGCTTGCCTAACCAGCAG
EPYC1_R:	ACATAACACACGCGTACCAAGGC
A1_F	GTTGGATGCACTAGTCACACGAGC
A2_F (EPYC1_Screen_pMJ016c_F)	GACGTTACAGCACACCCTTG
A1_R	GCACCAATCATGTCAAGCCT
E1_F	TCCTTCCGCACCAAAACATG
E2_F	CATAAGCTGTGAGCCGTTGA
E1_R	CAACTCAGTCAACGGCTCAC
E2_R (EPYC1_Screen_Gene_R)	ACAGTCGCATCAGAAAGGCT

Table S3. Raw qRT-PCR data

Gene	cDNA	Replica 1		Replica 2		Replica 3	
		Ct	Mean	Ct	Mean	Ct	Mean
Reference *RCK1/Cblp* (Cre06.g278222)	WT High CO_2	12.82	13.14	14.83	15.04	16.36	15.89
		13.3		15.35		15.6	
		13.31		14.93		15.7	
	epyc1 High CO_2	12.77	12.80	15.67	15.52	15.42	15.46
		12.82		15.41		15.38	
		12.81		15.48		15.58	
	WT Low CO_2	11.44	11.03	13.79	14.06	13.76	14.03
		11		14.48		14.53	
		10.65		13.92		13.81	
	epyc1 Low CO2	11.28	11.54	14.55	14.67		14.66
		11.53		14.63		14.67	
		11.82		14.84		14.65	
EPYC1	WT High CO_2	18.28	18.52	19.56	19.63	21.87	21.80
		18.5		19.64		21.8	
		18.78		19.68		21.73	
	epyc1 High CO_2	22.19	22.43	23.74	23.62	24.64	24.81
		21.98		23.66		25.14	
		23.13		23.46		24.66	
	WT Low CO_2	13.42	13.30	14.74	14.75	17.08	17.22
		13.25		14.75		17.52	
		13.24		14.75		17.06	
	epyc1 Low CO_2	22.64	22.47	24.28	24.22	24.87	24.82
		22.34		23.96		24.62	
		22.44		24.43		24.98	

Table S4. Raw O_2 evolution data

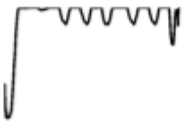
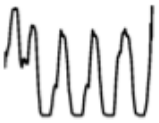
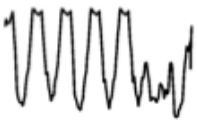
	Raw values (mV)										% of maximum													
	WT					*epyc1*					WT							*epyc1*						
	Low CO_2										Low CO_2													
uM *Ci	R1	R2	R3	R4	R5	R1	R2	R3	R4	R5	R1	R2	R3	R4	R5	mean	SD	R1	R2	R3	R4	R5	mean	SD
2.5			0.46			0.28	1.63	1.09		0.18			3.2			3.2		1.91		9.40		2.08	4.46	4.28
5	2.62	1.57	0.67	0.19	2.94	1.58		0.92	3.85	0.45	13.71	8.86	4.66	1.27	13.60	8.42	5.48	10.87	11.25	7.90		5.33	8.83	2.78
10	6.19	5.20	5.62	6.91	3.84	2.40	2.17	0.61		0.87	32.42	29.41	38.96	45.19	17.75	32.75	10.37	16.53		5.22		10.24	10.66	5.67
25	7.55	7.47	7.61	9.00	9.08	3.39		2.37	1.47	2.08	39.52	42.29	52.77	58.83	42.00	47.08	8.31	23.33	14.99	20.46	14.98	24.43	19.64	4.49
50	14.70	11.49		13.62	11.67	2.50		3.13	1.31	1.76	76.97	65.01	0.00	88.99	54.00	56.99	34.44	17.24		26.96	13.31	20.57	19.52	5.78
100	15.67	12.48	13.92	12.93	12.90	6.76	7.14	6.42	4.81	4.07	82.04	70.61	96.54	84.53	59.66	78.68	14.06	46.56		55.42	48.97	47.71	49.66	3.96
250		16.75	14.41	13.09	12.69	7.27	7.77	7.90	6.72	6.52		94.82	100	85.56	58.69	84.77	18.38	50.08	49.26	68.16	68.36	76.38	62.45	12.13
500	14.66		12.85	14.26	18.85	7.65	9.17	8.86	8.89	7.07	76.75		89.12	93.22	87.22	86.58	7.01	52.67	53.60	76.46	90.41	82.87	71.20	17.22
1000	19.10	14.88	13.49	15.30	21.61	14.52	14.50	11.59	8.79	8.53	100	84.21	93.63	100	100	95.57	6.92	100	63.23	100	89.43	100	90.53	15.93
2000		17.67	13.82	14.55	18.53		12.67		9.83	6.62		100	95.87	95.09	85.75	94.18	6.02		100		100	77.58	92.53	12.94
	High CO_2										High CO_2													
50	0.82	0.43	0.22			0.29	1.19	0.74			11.69	7.46	3.69			7.61	4.01	2.81	13.77	8.39			8.32	5.48
100	3.09	0.71	0.75			3.76	2.28	1.49			44.17	12.35	12.77			23.10	18.25	36.17	26.39	16.79			26.45	9.69
250	4.95	2.93	2.77			6.67	3.84	3.28			70.67	50.91	47.24			56.27	12.60	64.26	44.54	37.06			48.62	14.05
500	5.56	4.39	5.13			7.91	6.57	6.78			79.46	76.35	87.57			81.13	5.80	76.22	76.19	76.55			76.32	0.20
1000	7.00	5.37	5.85			9.88	7.26	8.10			100	93.53	100			97.84	3.73	95.11	84.16	91.48			90.25	5.58
2000	6.80	5.75	5.38			10.38	8.63	8.85			97.11	100	91.84			96.32	4.14	100	100	100			100	0.00

*Ci: Inorganic carbon

Table S5. Quantification of number of cells with multiple pyrenoids

WT Low CO_2				*epyc1* Low CO_2			
Image name	Total cells scored	Cells with >1 pyrenoid	Cells with 1 pyrenoid	Image name	Total cells scored	Cells with >1 pyrenoid	Cells with 1 pyrenoid
2015Feb23Freq25	30	1	12	2015Feb23Freq25	26	0	13
2015Feb23Freq24	24	0	11	2015Feb23Freq24	22	1	8
2015Feb23Freq23	26	0	7	2015Feb23Freq23	21	0	5
2015Feb23Freq22	30	0	13	2015Feb23Freq22	28	1	8
2015Feb23Freq21	17	0	8	2015Feb23Freq21	21	0	10
2015Feb23Freq20	22	0	5	2015Feb23Freq20	26	1	8
2015Feb23Freq19	25	0	9	2015Feb23Freq19	31	3	14
2015Feb23Freq18	26	2	10	2015Feb23Freq18	29	2	12
2015Feb23Freq17	25	1	7	2015Feb23Freq17	23	1	12
2015Feb23Freq16	28	0	12	2015Feb23Freq16	26	0	9
2015Feb23Freq15	23	0	11	2015Feb23Freq15	21	0	6
2015Feb23Freq14	20	0	10	2015Feb23Freq14	25	0	4
2015Feb23Freq13	27	0	15	2015Feb23Freq13	20	0	6
2015Feb23Freq12	23	0	8	2015Feb23Freq12	25	2	8
2015Feb23Freq11	22	1	12	2015Feb23Freq11	23	2	10
2015Feb23Freq10	24	0	10	2015Feb23Freq10	22	0	9
2015Feb23Freq9	22	0	9	2015Feb23Freq9	30	4	15
2015Feb23Freq8	26	0	10	2015Feb23Freq8	25	0	7
2015Feb23Freq7	24	0	6	2015Feb23Freq7	24	1	8
2015Feb23Freq6	30	0	17	2015Feb23Freq6	29	1	11
2015Feb23Freq5	22	0	7	2015Feb23Freq5	34	1	10
2015Feb23Freq4	21	1	6	2015Feb23Freq4	30	3	11
2015Feb23Freq3	24	0	12	2015Feb23Freq3	26	1	9
2015Feb23Freq2	37	1	12	2015Feb23Freq2	29	3	8
2015Feb23Freq1	25	1	13	2015Feb23Freq1	18	2	10
TOTAL	623	8	252	TOTAL	634	29	231
% of cells with multiple pyrenoids			3.2%	% of cells with multiple pyrenoids			12.6%

Table S6. Analysis of pyrenoid positive and pyrenoid negative algae for proteins with EPYC1-like physicochemical properties

		Number of proteins with…			Protein characteristics						
Species (Phylum)	Pyrenoid	…>=3 repeats with a 40-80aa repeat length…	…and a pI >8….	…and an oscillating disorder profile.	Protein ID(s)	Length	Repeat length	Repeat copy #	pI	Consensus repeat sequence	Disorder profile*
Chlamydomonas reinhardtii (Chlorophyta)	Y	18	8	1	Cre10.g436550 (EPYC1)	318	61	3.84	11.8	VTPSRSALPSNWKQELESLRSSSPAPASSAPAPARSSSASWRDAAPASSAPARSSSASKKA	
Thalassiosira pseudonana (Heterokontophyta)	Y	4	1	1	B8CF53_THAPS	376	53	6.21	9.1	LSSKPSSAPFVRSEKPSSAPSDSPSASVAPTLETSFSPSSSGQPSPMTSESPS	
Phaeodactylum tricornutum (Heterokontophyta)	Y	12	1	1	B7GDW7_PHATC	380	46	7.17	9.9	TGPSMTGPSDSDDRRLRSPSSTGPSLTGPSMTGPSATGPSMTGPSM	
Emiliania huxleyi (Haptophyta)	Y	99	10	2	R1G412_EMIHU	353	70	4.10	12.1	PYLPISPARLARGSTSPHLSPSLPISPHISRTARSRFHIAPSLPISPHISPTAPHGFHEAPHLPISPHLS	
					R1D601_EMIHU	255	60	3.70	10.1	WTAADDALVKAGQEAGESWVDIAKRLPGRSADSVKSRSNRLKRQPDTSVKHEPVKRELVR	
Micromonas pusilla (Chlorophyta)	Y	6	0	0							
Chlorella variabilis (Chlorophyta)	Y	3	2	1	E1ZFJ0_CHLVA	321	70	3.81	9.5	LGLGAASLLAAQNAEAAQQVADLAAGDNRFGTIAFLALVPVVGWVLFNILGPLQNQLDAMDTKKRSVA	
Ostreococcus tauri (Chlorophyta)	Y/N†	3	3	2	A0A096PAN3_OSTTA	407	63	5.02	11.1	MAASKLGSKNASTRPTVGSTLDASALTPPSLRFTTENNIHSVPTAFGVADRPASRRVLRREDA	
					A0A090M8K8_OSTTA	470	63	6.02	11.2	MAASKLGSKNASTRPTVGSTLDASALTPPSLRFTTENNIHSVPTAFGVADRPASRRVLRREDA	
Chlorella prototheocoides (Chlorophyta)	N	1	0	0							
Cyanidioschyzon merolae (Rhodophyta)	N	0	0	0							
Galdieria sulphuraria (Rhodophyta)	N	2	0	0							
Nannochloropsis gaditana (Heterokontophyta)	N	3	0	0							

*Disordered profiles are a plot of disorder propensity (y axis; 0-1; 0 = ordered, 1 = disordered) against amino acid number (x-axis; 0-437). All profiles are on the same scale. †TEM images of *Ostreococcus tauri* show a singular starch deposit typical of a pyrenoid, however a Rubisco matrix has yet to be confirmed (67).